THE UNIVERSITY
of EDINBURGH

THE UNIVERSITY *of* EDINBURGH

Expanding the CRISPR toolbox for use within *Bacillus subtilis*

Marcus Alexander Price

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2019

Abstract

DNA editing is a vital tool in the development of biological systems for both research and commercial applications. Novel enabling tools accelerate strain engineering for the study of cellular mechanisms or production of small molecules and proteins. CRISPR (Clustered Regularly Interspersed Short Palindromic Repeats) technologies, where small RNA molecules (gRNA) direct Cas (CRISPR-associated) proteins to DNA in a highly accurate, sequence-dependent manner have increased the rate at which DNA modifications can be made.

We developed and expanded the CRISPR toolbox for recombination and deaminase-guided genome editing, protein engineering and transcriptional regulation within the industrial workhorse *Bacillus subtilis*.

A co-transformational system, consisting of a single plasmid for nuclease and gRNA expression and a linear donor DNA (dDNA) was established for use with both *Streptococcus pyogenes* Cas9, the most commonly utilised CRISPR nuclease, as well as the potentially highly commercially relevant nuclease, MAD7, also known as *Eubacterium rectale* Cas12a. Editing efficiencies of ≥83% were observed for both nucleases.

Using our CRISPR-Cas9 genome editing tool, a novel variant of the commercially relevant protein, subtilisin E, was engineered exhibiting an increase in both thermostability and proteolytic efficiency. Two systems for transcriptional down-regulation (CRISPRi) were demonstrated, including the first reported catalytically inactive variant of MAD7. Finally, Cas9 was further modified to incorporate fusions with DNA deaminases allowing the first reported example of CRISPR targeted deaminase base editing within *B. subtilis*.

Lay Summary

Over the years, bacteria have got quite a bad reputation within the general public for the spread of illness in people and animals. Not all bacteria are bad though. For example, *Bacillus subtilis* has been designated as safe for decades and is found widely in the soil, as well as in some foods.

Another great use for *B. subtilis* is that it has an incredible ability to release enzymes into the environment around it. This ability has been used for a very long time to make enzymes that we use every day in our society. Uses include enzymes that help clean clothes in biological washing powder, and to help break down corn to make bioethanol which is used in fuel for cars. We can make these enzymes faster and develop new enzymes with better properties, for example to work well in a more diverse and potentially harsh environment. To make these changes though we need the ability to change the genetic code, DNA, within the bacteria.

In 2012, a type of immune system used by some bacteria known as CRISPR-Cas9 was modified to quickly and highly accurately change DNA. CRISPR-Cas9 causes a cut in DNA which has to be repaired or prevented for the cell to survive. It is in the repair of this cut that we can introduce changes. We can remove DNA, add new DNA and make precise changes without changing the total amount of DNA.

The work in this thesis adapts some of the tools to change DNA in other bacteria, and more complex organisms, for use in *B. subtilis*. We successfully show the basic CRISPR-Cas9 system to be just as effective in *B. subtilis,* as well as further developing a tool which is free to use within industry to maximise the impact of this technology. We used these tools to develop a new form of an enzyme, which is used within washing powder, and works 46% faster than the unchanged enzyme.

We hope the tools developed here are used broadly within the scientific community to increase the speed and accuracy by which new, better products can be created. This will hopefully broaden the uses for enzymes and further reduce our society’s reliance on oil-based products.

Declaration

I certify that this thesis contains my own work, except where acknowledged, and that no part of this material has previously been submitted for a qualification at this or any other university. The publications in Appendix E and Appendix G is my own work in collaboration with the other indicated authors whose individual contributions are outlined within the publication.

Marcus Alexander Price

03 March 2020

Acknowledgements

Firstly, I'd like to thank Prof. Susan Rosser for giving me the opportunity to do this PhD and for providing me with support, encouragement and advice whenever I needed it. Susan's whole group were incredibly welcoming and helpful to me and I never felt like an outsider during my weekly visits for the lab meeting. I would also like to thank Ingenza for giving me such a fantastic place to work, supporting me throughout my PhD and allowing me to develop friendships which will last a lifetime.

I'd like to say a really special thank you to Dr Rita Cruz who has supervised, mentored, collaborated and supported me with endless patience and friendship. Rita has massively improved my scientific ability, self-confidence and professionalism (not to mention my preference in colours)! She is in my opinion the blueprint of a great scientist and person. I feel very fortunate to have learned from her and to have her as a close friend!

In my opinion, one of the most important things you need during a PhD is support from the people around you at work. Ingenza provided me with a brilliant environment and colleagues with which to work. Thank you in particular to Dr Franck Escalettes, Dr Scott Baxter and Dr Leonardo Magneschi for advice, technical discussions and great collaborations. Thank you very much to Dr Ian Fotheringham, Dr David McElroy and Jaymin Amin for enabling my PhD to go ahead at Ingenza and supporting me throughout. To my lab sister Annemette Kjeldsen, who joined me in China and Ingenza, before and during her own PhD, thank you for all the chats, rugby, delicious yogurt drinks, whisky and watermelons. Helping to mentor Louis Marlow throughout his masters project was a joy and it was really nice to see him at Ingenza when I came back from China! Thank you also to Pamela Ford and Gail Campbell for making so many plates for me and being so patient with the odd late order! A big thank you to all other Ingenzers, past and present, for countless laughs, questions, advice and, most importantly, cake!

Without the funding support from The Industrial Biotechnology Innovation Centre (IBioIC) this PhD would never have happened so for that I thank them and in particular Dr Ian Archer for advice and encouragement. I also had the fantastic chance of spending three months in Tianjin, China, with Prof. Changhao Bi. Thank you to the Newton Fund and British Council for funding this placement. Thank you Prof. Bi, Siwei Li and Dongdong Zhao for welcoming me into their lab and country,

as well as collaborating with me for chapter 6 of this thesis and on papers for which the manuscripts are in preparation or under review.

To all my close friends who have been relentless on their thoughtfulness, advice and proof-reading skills, in person and via WhatsApp chats, thanks a million! In particular, James Kinkead, Daniel Cooney, John Munro, Cameron Beattie, Lewis Fenton and Callum Campbell – I'll always cherish the DNA of our friendship! Thank you also to Sarah Young for answering all my questions about homologous recombination!

It's important to be able to switch off from work. To that end I would like to thank the Capital Concert Band, in particular Karyn McGhee, Gemma Milligan, Kirsty Dingwall, William Robertson, Michael Garner, Neil Mackenzie, Susanne Jackson and Kevin Lamb for musically resetting my mind once a week. Furthermore, I'd like to thank James Kinkead, Fraser Brown, Sharon Versteeg and Sally-Anne Hepburn for getting me out into the hills regularly on my bike to shred some gnar! Also, to the entire Scottish and Edinburgh rugby teams for giving me stuff to shout at and about, as well as making me accustomed to the feeling of disappointment – all too common in science.

The support of my parents, sisters and wider family has been brilliant, even though they maybe don't get what I'm saying all the time – it's difficult translating science into German sometimes! I wouldn't be the person I am today without them all!

Lastly, and by far and away most importantly, I'd like to thank my amazing wife, Fiona. Every part of this work is dedicated to her. Fiona has been so rock solid in her support of me throughout this PhD that I cannot imagine doing any of this without her. Simple things in life can take on an enormity when people are stressed and so Fiona doing these, and thousands of other things (like planning our wedding!) has helped my mental health massively. You're my best friend and partner in crime! Thank you!

Contents

List of Publications

Paper Title	Journal	Contribution	Status	Ref.
CRISPR-Cas9 *In Situ* engineering of subtilisin E in *Bacillus subtilis*	PLoS One	First author	Published – 2019	[1]
Expanding and understanding the CRISPR toolbox for *Bacillus subtilis* with MAD7 and dMAD7	Biotechnology and Bioengineering	First author	Published – 2020	[2]
Glycosylase base editor (GBE) enables specific C to A and C to G genomic conversion	Nature Biotechnology	Sixth author	Accepted upon revision	
Helicase-AID: a novel molecular device for base editing at random genomic loci		Sixth author	Redrafted for resubmission	
CRISPR-enabled deaminase base editing tools for utilisation in *Bacillus subtilis*		Co-first author	Manuscript in preparation	

List of Figures

List of Tables

List of Equations

List of Abbreviations

°C – degrees Celsius

ABC – ATP-binding cassette

ABE – deoxyadenosine base editing

AID – activation-induced cytidine deaminase

Amp – ampicillin

AP – apurinic/apyrimidinic site

ATP – adenosine triphosphate

A.U. – arbitrary units

BER – base excision repair

bp – base pairs (can be combined with metric SI prefixes)

CBE – deoxycytosine base editing

Cm – chloramphenicol

cPCR – colony PCR

CRISPR – clustered regularly interspersed palindromic repeats

CRISPRa – CRISPR activation

CRISPRi – CRISPR interference

Cas – CRISPR associated

cDNA – complementary DNA

CFU – colony forming unit

CGQ – Cell growth quantifier

CIP – calf-intestinal phosphatase

Cpf – CRISPR from *Prevotella* and *Francisella*

C_q – quantification cycle

crRNA – CRISPR RNA

dATP – deoxyadenosine triphosphate

dCTP – deoxycytidine triphosphate

dDNA – donor DNA

DDT – dithiothreitol

dGTP – deoxyguanosine triphosphate

DNA – deoxyribonucleic acid

DR – direct repeat

DSB – double-strand break

dNTPs – Deoxyribonucleotide mix of bases dATP, dTTP, dGTP and dCTP

dsDNA – double-stranded DNA

dTTP – deoxythymidine triphosphate

dUTP – deoxyuridine triphosphate

EDTA – ethylenediaminetetraacetic acid

EFSA – European Food Safety Authority

F – farad (can be combined with metric SI prefixes)

FDA – Food and Drug Administration

g – gram (can be combined with metric SI prefixes)

g – gravity

gDNA – genomic DNA

GFP – green fluorescent protein

GRAS – generally regarded as safe

gRNA – guide RNA

HDR – homology directed repair

HNH – Histidine-Asparagine-Histidine

HR – homologous recombination

HS – high specificity

IPTG – isopropyl β-D-1-thiogalactopyranoside

Kan – kanamycin

KI – knock-in

KO – knock-out

L – litre (can be combined with metric SI prefixes)

LB – lysogeny broth

LIC- ligation independent cloning

m – meter (can be combined with metric SI prefixes)

MCS – multiple cloning site

min – minute

mol – mole (can be combined with metric SI prefixes)

MTP – microtitre plate

NHEJ – non-homologous end joining

nt – nucleotide

NUC – nuclease domain

OAPS – oligonucleotide annealing based promoter shuffling

OD_{600} – optical density at 600 nm

OE-PCR – overlap extension PCR

PAGE – polyacrylamide gel electrophoresis

PAM – protospacer adjacent motif

PCR – polymerase chain reaction

PEG – polyethylene glycol

QPS – qualified presumption of safety

R&D – research and development

RBS – ribosome binding site

rcf – relative centrifugal force

REC – recognition domain

RNA – ribonucleic acid

RNAP – RNA polymerase

rpm – revolutions per minute

RT qRT-PCR – real time quantitative reverse transcription PCR

SD – standard deviation

SDS – sodium dodecyl sulphate

SEC – size exclusion chromatography

s - second

sgRNA – single guide RNA

SM – starvation medium

Spec – spectinomycin

SRP – signal recognition particle

SSB – single strand binding protein

ssDNA – single-stranded DNA

TAE – tris-acetate-EDTA

tracrRNA - *trans*-activating CRISPR RNA

TALENs – transcription activator-like effector nucleases

Tat – twin arginine translocation

T_m – melting temperature

UV – ultra violet

V – volts (can be combined with metric SI prefixes)

v/v – volume per volume

WT – wild-type

w/v – weight per volume

ZFNs – zinc finger nucleases

Chapter 1 – Introduction

1.1 Context and motivation

Historically, production of chemicals and energy has relied on fossil fuel-based approaches. In the current context, there is an urgent need to move away from this model and towards a set of processes with a lower environmental impact. One promising approach is to replace energy-intensive chemistry with natural processes. Biotechnological applications, such as industrial enzyme production or riboflavin (vitamin B2) secretion by *Bacillus subtilis*, have the capacity to reduce significantly environmental impact in comparison to their industrial chemistry equivalents[3]. However, to reach a point of commercial viability, bioprocesses within a host must be optimised to increase productivity rates. Therefore, it is important to increase the speed and accuracy with which genetic modifications can be made.

Since the advent of simple, targeted genome editing in 2012 with the exploitation of the Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR) and CRISPR-Associated (Cas) proteins, many areas of the biological sciences involving genomic modifications have been revolutionised. Experiments that would previously have taken considerable time and resources with transcription activator-like effector nucleases (TALENs) or zinc finger nucleases (ZFNs) can be reduced to a fraction of the time and cost[4]. Where previously expensive, complex and laborious methods were required to make relatively small changes, now the scientific community has a tool for rapid, relatively cheap and simple changes from the single base pair alteration, to the modification of an entire genome through multiplexed (multiple targets) genome deletions[5].

The bacterial industrial workhorse *B. subtilis* is used globally for the large-scale manufacturing of enzymes used in diverse applications such as the food, textile, paper, detergent, pharmaceuticals and bioethanol production industries, as well as bulk chemical production[3,6]. It is found naturally in the soil, and is also a model organism for Gram-positive bacteria. The development of CRISPR tools for use within *B. subtilis*, building on the well-established protocols for genetically modifying this host, has the potential to increase the rate by which new bioprocesses are designed, constructed, exemplified and optimised.

1.2 *Bacillus subtilis*: industrial workhorse and model organism

B. subtilis is one of the most intensively studied microorganisms. Its lack of endotoxin production has enabled its qualification as “generally regarded as safe” (GRAS) from the United States Food and Drug Administration (FDA), and the similar QPS (qualified presumption of safety) by the European Food Safety Authority (EFSA). It has been used for centuries within the Japanese food, including in “nattō” a breakfast food produced from soybeans fermented with *B. subtilis* derived from rice straw[7]. This, coupled with its well understood, inducible genetic competence pathway has made it one of the primary industrially utilised bacterial species[3,8].

Since Bacilli are capable of 20-25 g/L secretion of proteins[3,9,10], their use as enzyme production hosts has been extensive, with production of ~60% of commercially available enzymes[10] within a global enzyme market with a value of $5.5 billion in 2018, and predicted to reach $7 billion by 2023[11]. The enzymes produced vary in function with some examples being amylases, amylomaltase, cellulases, proteases and xylanases[3]. While *B. subtilis* is the model for other Bacilli, notable alternative production hosts have preferential characteristics, depending on the target enzyme being produced. *Bacillus licheniformis* is widely used for alkaline serine protease production, can grow anaerobically[8] and its optimal growth temperature is ~46 °C[12]. *Bacillus amyloliquefaciens* secretes its native amylase at high levels, is used for starch degradation in the beverage industry[9], and has also been shown to improve plant tolerance to abiotic stress such as elevated salt concentrations[13].

Bacilli are not only used for the production of proteins. They are also used to produce fine chemicals such as nucleotides, which are used in the food industry as the umami flavour enhancer and had a global demand of 23,000 tonnes in 2009[3]. *B. subtilis* has been modified to achieve >14 g/L production of adenosine[14], while *B. amyloliquefaciens* has been shown to achieve secretion of 10 g/L inosine[3]. Vitamin B2, also known as riboflavin, has an annual global demand of 4000 tonnes and has been produced from *B. subtilis* (>15 g/L)[15].

1.2.1 Natural competence and genetic manipulation of *B. subtilis*

The increased knowledge of the genetic competency mechanisms, and development of natural competence induction by Anagnostopoulous and Spizizen[16], has made *B. subtilis* the model organism for Gram-positive bacteria. The development of online tools, such as SubtiWiki, has facilitated greater understanding through improved communication of data[17].

B. subtilis has the capacity to differentiate into several physiological states, depending on the environmental conditions. When the vegetative cell enters stationary growth phase, subpopulations become motile, form biofilms, secrete degradative enzymes and antibiotics, become naturally competent and lastly sporulate for long-term survival[18]. The main factor in *B. subtilis* differentiation is the level of phosphorylated Spo0A. This acts as a global regulator which, along with AbrB, CodY, Rok and the DegS-DegU two-component system, act to regulate tightly the transcription levels of the competency master regulator, ComK[18]. Such tight regulation leads to natural competence only being achieved by ~10% of cells of *B. subtilis* within a growth culture[19] and even fewer *B. licheniformis* cells[20]. Once at sufficient levels within the cell, ComK causes the cell to enter a state of competency through the activation of expression for ~100 genes involved in DNA uptake, repair, binding and recombination[21]. Double-stranded DNA (dsDNA) binds to a receptor (ComG) via a pseudopilus, then is cleaved by a membrane-bound nuclease and internalised as a linear single-stranded DNA (ssDNA) molecule by ComEA[22]. Strains of *B. subtilis*[18] and *B. licheniformis*[20] have been modified to incorporate an inducible expression cassette for ComK to allow the induction of natural competence, as an alternative to the starvation-based approaches used in the absence of such expression[16,23,24].

The ssDNA is protected from exonucleases within the cell following the binding of RecN to the 3´-OH[25]. DprA, SsbA and SsbB also bind to the internalised ssDNA. DprA has been found to be more important during the transformation of replicative plasmids due to the increased stabilisation required to allow ssDNA strand annealing and subsequent DNA synthesis[26]. Furthermore, DprA facilitates the replacement of SsbA and SsbB with RecA, which goes on to be the major component enabling homologous recombination (HR) - discussed in detail in the context of double-strand break (DSB) repair within section 1.4.2[18].

This efficient competency and HR capacity has led to several tools being developed for the genetic manipulation of *B. subtilis* to construct production strains without antibiotic resistance genes. Counterselection mechanisms are such an example. Here, an inducible expression cassette for a toxic gene such as the *Escherichia coli* toxin gene, *mazF*, can be incorporated alongside an antibiotic gene. The antibiotic gene enables selection of colonies which have undergone successful integration events. Induction of *mazF* expression produces a toxic endoribonuclease which cleaves free mRNA at ACA sequences. Thus, for the *B. subtilis* host to survive, *mazF* must be removed. This occurs between homology sites generally flanking *mazF*, as well as the antibiotic resistance gene. Thereby, *mazF* and the antibiotic resistance gene are removed and an engineered host constructed[3,27]. Other counterselectable markers include the *upp* gene in the presence of 5-fluorouracil (requiring inactivation of the native *upp* gene)[28], the *blaI* cassette in a lysine auxotroph[29], and the heat-inducible hen egg white lysozyme gene[30].

Following transformation by HR, removal of the toxin and antibiotic resistance genetic elements can be further boosted through the use of a recombinase enzyme[31]. The bacteriophage P1 Cre/*loxP* site-specific recombination system utilises the Cre recombinase to catalyse the reciprocal, site-specific recombination between two *loxP* sites, excising the DNA between the sites (such as an antibiotic resistance gene) and leaving a single *loxP* site as a scar[31]. However, this system was not readily compatible with multiple rounds of editing due the presence of the *loxP* scar interfering with the subsequent round of editing by removing the incorrect section of DNA between the initial *loxP* scar and the subsequent *loxP* sites. The development of the *lox66* and *lox71* mutant sites addressed this concern as they recombine to leave the 34 bp *lox72* scar which has a greatly reduced binding affinity for Cre and thus does not interfere with subsequent rounds of genomic modifications[32]. While counterselectable systems may require a specific background, or result in spontaneous resistance to the counterselection mechanism, Cre/*loxP* editing does not require any specific background in the cell. That being said, Cre/*loxP* results in strains containing scars at each editing region, while the *mazF* counterselection mechanisms is scarless.

1.2.2 Bacilli as protein production hosts

B. subtilis, as well as other Bacilli, are capable of secreting proteins to a high titre within industrial fermentation processes (20-25 g/L)[3,9]. As such, protein secretion pathways have been highly analysed and their function is now well understood[3,33]. The primary secretory pathway utilised, Sec, is discussed below. The other known secretory pathways, the twin arginine translocation (Tat) pathways, ATP-binding cassette (ABC) transporters and the pseudopilin export pathway are characterised[33].

1.2.2.1 Protein secretion via the Sec pathway

The Sec pathway is the primary protein secretory pathway used within *B. subtilis*[33]. Unfolded protein polypeptides are targeted to the cytoplasmic membrane by signal peptides found at the N-terminus of the eventual final protein. These consist of a positively charged N-terminal region, a hydrophobic region and a polar C-terminal region[10]. Following ribosomal translation, the peptides are kept in their unfolded, translocation-competent state by the signal recognition particle (SRP) and FtsY chaperones[34]. The Sec apparatus, which consists of the SecA ATPase motor (activated by interactions with the signal peptide[33]), and the membrane channel proteins SecY, SecE, SecG (SecYEG) and SecDF, can then be utilised to pass the peptide through the cytoplasmic membrane. The exact function of SecDF is still unknown, however it is required for efficient protein secretion, particularly in high-scale protein production hosts (≈1.3 g/L) and under cold conditions (15 °C)[35]. Once the peptide has passed though the SecYEG channel, the signal peptide is cleaved at the C-terminal region and degraded by signal peptidases, such as SppA, which is important as undegraded signal peptides can inhibit translocation of other proteins[33].

Once translocated, the protein is correctly folded in the extracytoplasmic space. The extracellular chaperone, PrsA facilitates protein folding in the majority of cases[36], but this process can also be assisted by thiol-disulphide oxidoreductases BdbB, BdbC and BdbD[37]. Some *B. subtilis* proteins, such as the serine protease subtilisin E (encoded by *aprE*) have a further propeptide which assists in processing the enzyme into its final form, before autoprocessing the removal of the propeptide[38]. Furthermore, metal ions such as Fe^{3+}, Mg^{2+} and, in the case of Subtilisin E, Ca^{2+}, aid folding and stabilise the final active enzyme. Lastly, the

membrane-bound proteases HtrA and HtrB, as well as the cell wall associated protease WprA act as “quality control” proteases, ensuring there are no blockages in the secretion system, before the final secreted protein is released into the extracellular environment.

1.2.2.2 Other protein secretion pathways in *B. subtilis*

The alternative secretory pathways within *B. subtilis* are the Tat pathway, ABC transporters and very rarely the pseudopilin export pathway. Secreted proteins utilising the Tat pathway have a highly conserved twin-arginine motif within their signal peptides and are transported across the cell membrane in their folded state and can include tightly folded proteins and multimeric proteins, such as the phosphodiesterase, PhoD, involved in degradation of teichoic acid within the cell wall during phosphate starvation[3,33]. ABC transporters function as multidomain permeases to import or export various ions, amino acids peptides antibiotics or proteins across the cell membrane in prokaryotes and eukaryotes. In *B. subtilis*, the signal peptide structure for proteins destined for ABC transporters, such as lantibiotics, is significantly different to that utilised by the Sec or Tat pathways in that it lacks the hydrophobic region[33]. Lastly, the pseudopilin export pathway is involved in the transport of proteins required for the development of natural genetic competency[33]. These proteins remain in the extracytoplasmic space between the membrane and the cell wall to fulfil their function[33].

1.2.2.3 Alteration of the extracellular proteome of *B. subtilis* to maximise protein yields

Within industrial fermentation processes, to maximise extracellular levels of the target protein, the extracellular proteome of *B. subtilis* can be altered. Deletion of extracellular proteases can substantially increase the level of heterologous protein production. *B. subtilis* secretes at least eight extracellular proteases through the Sec pathway. These are, the previously mentioned and model for protein engineering[39], alkaline serine protease, subtilisin E (AprE), as well as neutral protease (NprE), minor protease (Epr), bacillopeptidase F (Bpr), metalloproteases (Mpr), cell wall associated WprA, Vpr and NprB proteases. Of these, 95% of the extracellular proteolytic activity is attributed to AprE and NprE[3]. Knock-outs of these proteases

have been constructed, producing strains without antibiotic resistance markers or extracellular proteases as a platform for heterologous protein production[40,41].

Further deletions of major extracellular proteins, such as the starch-degrading α-Amylase (encoded by *amyE*), can also aid in downstream processing by limiting the level of contaminants required for removal during target protein purification. Lastly, deletion of the surfactin synthase protein SrfAC has been found to reduce the level of foaming during large-scale fermentation, thus reducing product loss[41].

1.3 CRISPR-Cas mechanisms and their utilisation for genome engineering

1.3.1 An immune system for bacteria and archaea

CRISPR-Cas systems are natively found in 50% of analysed bacterial genomes and 87% of analysed archaea genomes[42]. These act in a similar manner to an immune response when the host is targeted by, bacteriophages, or plasmid DNA invades the cell[42,43]. CRISRP-Cas systems have been split into six distinct types (I-VI), with each type having its own Cas proteins associating to CRISPR RNA (crRNA) to induce DNA interference[42,44]. The most widely studied type of CRISPR system is type II, in which a single Cas9 protein performs a wide variety of roles in contrast to type I and III systems where a complex of Cas proteins perform the same role. The type V system, comprising Cpf1 nucleases (also referred to as Cas12a), is also a multipurpose enzyme, however it differs significantly from Cas9 in that it is guided by a single RNA molecule, as opposed to two, and cleaves DNA in a staggered manner, as opposed to a blunt DSB[44,45].

In the type II CRISPR-Cas9 system, the natural process occurs in three stages. Firstly, the cell acknowledges the invading viral or plasmid DNA and incorporates short sections (protospacers) into the CRISPR loci. These are then transcribed into precursor CRISPR RNA (pre-crRNA) consisting of a 20-25 bp targeting region, and a stretch of 12 bp which is constant in all pre-crRNA[43,46]. Pre-crRNA are matured to crRNA by association with a *trans*-activating CRISPR RNA (tracrRNA)[42,43]. These mature crRNA-tracrRNA complexes allow binding of the foreign DNA by the Cas9 protein. The invading DNA is then cleaved 3 bp upstream of the protospacer adjacent motif (PAM) which is required for Cas9 activity[42,43]. The

newly acquired protospacer is maintained in the CRISPR loci, and upon invasion of the same stretch of DNA, transcription of the pre-crRNA and expression of Cas9 removes the invading DNA[42,47,48].

1.3.2 Characterisation of CRISPR-Cas9 for its use within genomic engineering

Due to its simplicity the type II system, utilising Cas9 as described here, has become the most widely utilised[42]. The potential of CRISPR-Cas9 for highly targeted genomic modifications through DSB repair mechanisms was identified by Jinek *et al*.[43]. The relatively common PAM motif (5´-NGG-3´) required by *Streptococcus pyogenes* Cas9 (henceforth, Cas9 refers to that from *S. pyogenes* unless otherwise stated) in all organisms led to this becoming the system of choice for genome editing, with only one heterologous gene required for highly efficient genome editing (*cas9*). Type II systems contain tracrRNA which associates to, and aids in the maturation of, the guiding crRNA[42,46]. Development of the crRNA-tracrRNA into a single guide RNA (sgRNA), joined by a linker loop, simplified the system further, and was found to be just as or more efficient than when separated[43,47]. sgRNAs have since been further developed, for example to include a secondary structure preventing hybridisation at an incorrect target, in an attempt to limit the off-target effects observed with CRISPR systems in eukaryotes[49].

1.3.2.1 Target interrogation and cleavage by Cas9

Cas9 is a bilobe endonuclease composed of the alpha-helical recognition (REC) lobe and the nuclease (NUC) lobe which contains both the HNH (Histidine-Asparagine-Histidine, targeting the DNA strand complementary to the gRNA) and RuvC (named after an endonuclease domain found in *E. coli* RuvC) nuclease domains[42,47]. Crystal structure of the isolated, apo-Cas9[50], shows the apo form to be substantially altered at the PAM recognition site following binding to the crRNA-tracrRNA/sgRNA complex, yielding holo-Cas9 (Figure 1.1)[46]. This inactive state of the PAM recognition region supported the findings of Jinek *et al*. in which Cas9 was found to be inactive as a nuclease without the presence of a gRNA[43].

Once bound to the targeting RNA, Cas9 interrogates the DNA for the presence of a PAM site. Following identification of such a site, the length of time the

nuclease interrogates that site is determined by the homology to the targeting RNA. The 10-12 nt region directly upstream of the PAM site, within the targeting region, is known as the seed region (Figure 1.1). Any mutations within this seed region have a highly deleterious effect on Cas9 activity[43,51], while mismatches within the PAM-distal end of the gRNA is tolerated to a greater degree[52].

Once the Cas9-RNA complex has found the target site, the Cas9 relaxes the DNA at the target site, known as the R-loop. This enables gRNA strand invasion and the formation of a DNA-RNA duplex, through Watson-Crick base pairing, and a single DNA strand (Figure 1.1)[53]. Once the R-loop has fully unwound to allow interrogation of the full protospacer, provided there are no mismatches within the seed region and no more than 3 mismatches, the DNA is cut by the HNH and RuvC nuclease domains.

Holo-Cas9 is only fully active when the full gRNA hybridises to the DNA enabling a conformational change in the HNH domain, which in turn promotes a change in the RuvC domain, in each case bringing the target DNA into the catalytic site of the domain[53]. The HNH domain cleaves the DNA strand to which the gRNA has hybridised, while the RuvC domain cleaves the complementary strand[43,54]. The cut induced by Cas9 is blunt and takes place 3 bp upstream of the PAM site, within the seed region (Figure 1.1). Following DNA cleavage, Cas9 remains bound to the cleaved DNA until displaced by other cellular factors[55].

The Cas9 derived DSB must be repaired or inhibited for the host to survive. It is this requirement that provides the CRISPR-Cas systems with their unparalleled genome editing efficiency when compared to established editing techniques in *B. subtilis* (section 1.2.1), and when compared to the complexity of changing the target location for other targeted nuclease editing techniques, such as TALENs or ZFNs, where complex protein modifications are required for accurate DNA cleavage[4]. The DSB can be prevented by HR of a donor DNA (dDNA), or repaired by cellular DNA repair mechanism, of which the most common are the error-prone non-homologous end joining (NHEJ) or the more controlled homology directed repair (HDR) when a dDNA template is present – generally a polymerase chain reaction (PCR) product, or a plasmid borne template containing the intended modification as well as an alteration of the PAM site to prevent further cutting (DNA break repair mechanisms are discussed in detail in section 1.4)[47].

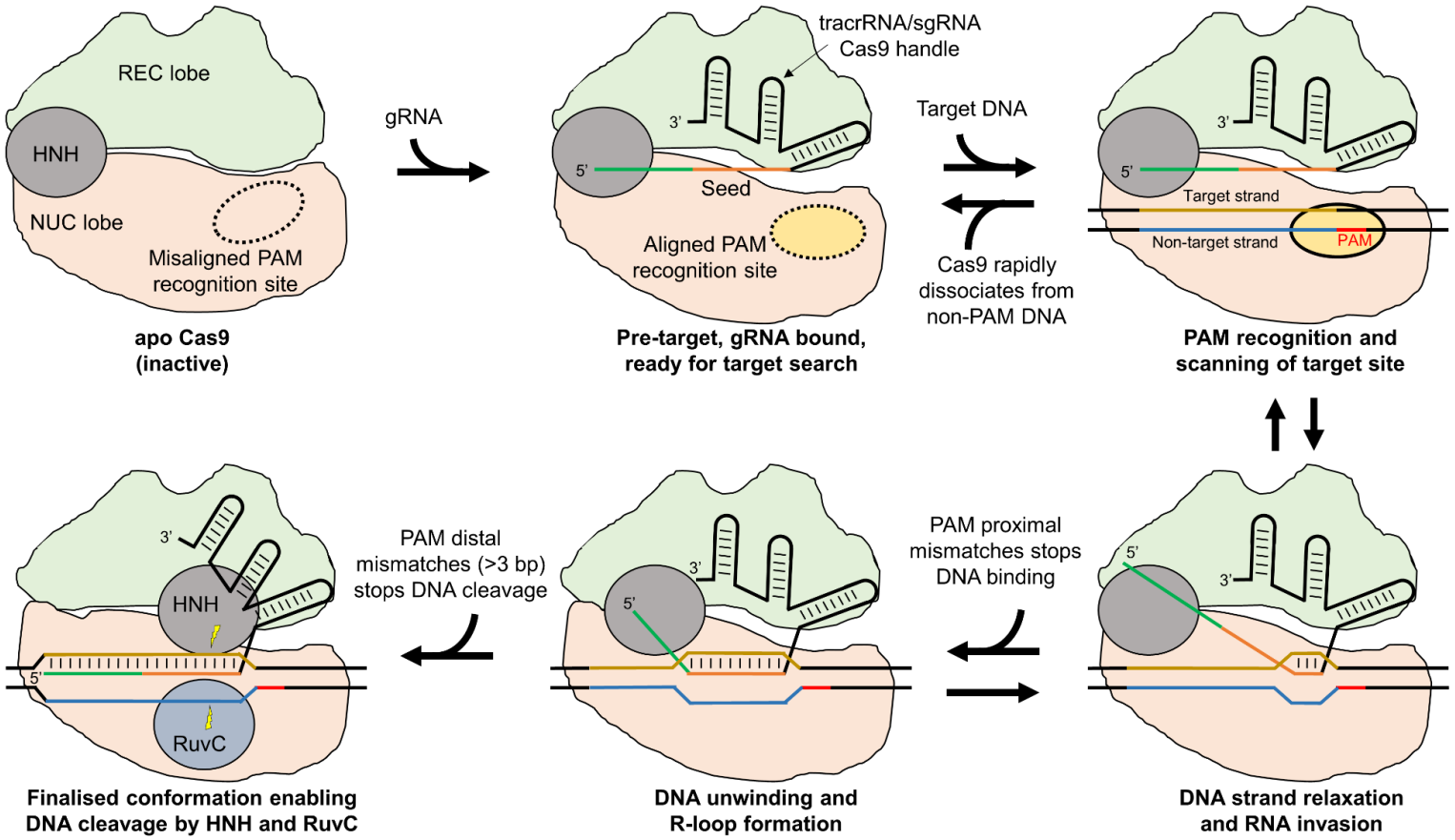


Figure 1.1 – Schematic of the mechanism by which CRISPR-Cas9 associates to, and cleaves its DNA target.

Cas9 undergoes a conformational change, aligning the PAM interacting region correctly, following the binding of guide RNA (shown here as a combined single guide RNA (sgRNA), but this can also be individual crRNA-tracrRNA components). Abbreviations: NUC, nuclease lobe; PAM, protospacer adjacent motif; REC, recognition lobe.

Cas9 has been shown to have two crucial amino acids for its catalytic activity (Asp10 and His840)[43]. It was established that replacing one or both of these with an alanine residue would lead to a partially or fully inactive variant respectively, while maintaining its capacity for binding DNA[43,47]. Catalytically altered versions of Cas9 are known as nickase Cas9 (nCas9 – the DNA strand specific D10A or H840A mutations preventing cleavage of a single DNA strand though inactivation of the RuvC or HNH nuclease domains respectively) and deactivated/dead Cas9 (dCas9 – both D10A and H840A mutations are present). These have become widely used since their development with just some examples being nCas9 used to increase genome editing accuracy in mammalian hosts through the limitation of off-target DSBs (off-target single strand nicks are efficiently repaired while editing only takes place when two nicks, from different gRNA, are in close proximity to create a DSB)[56], and enabling selection of strains which have undergone successful deaminase base editing (section 1.3.5) by cleavage of the non-edited strand[54,57]. dCas9 has also been utilised for deaminase base editing[54,57], but is perhaps most broadly utilised for transcriptional regulation where, in bacteria, it sterically hinders the binding or progression or RNA polymerase (RNAP). Alternatively, when fused to an effector protein, RNAP can be recruited to a specific site at a greater rate, resulting in an increase in transcription of the targeted gene[58–61].

1.3.3 CRISPR nucleases with alternative PAM site, cleavage and guide RNA characteristics

Cas9 has been engineered to accept alternative PAM sites[62]. Although the NGG PAM sequence occurs relatively often in organisms with a medium to high GC content, less options for editing targets are available in AT-rich genomic regions. As a result, an NGG PAM site could be prohibitively distant from the point at which editing is desired. Having several Cas9 variants accepting alternative PAM sites allows closer, targeted editing to occur[62]. As well as variants of *S. pyogenes* Cas9, Cas9 nucleases from other microorganisms allow targeting of different PAM sites (Table 1.1). Furthermore, the utilisation of the type V CRISPR-Cpf1 nucleases, including the recently released commercially relevant MAD7 nuclease, offer alternative PAM sites in AT rich regions (Table 1.1).

Table 1.1 - Cas9/Cpf1 proteins and corresponding PAM sites. These have been shown to be functional for genome editing in eukaryotic cells. N = A, T, G or C.

Cas Protein	PAM Sequence (5´-3´)	Origin Organism	Exemplified use in prokaryotes?
Cas9	NGG	*Streptococcus pyogenes*	Yes[43,63]
VRQR-Cas9	NGA	Engineered from Cas9.	No[64]
VRER-Cas9	NGCG	Engineered from Cas9.	No[64]
Nm_Cas9	N_4GMTT (M = A or C)	*Neisseria meningitides*	No[64]
Sa_Cas9	N_2GRRT (R = A or G)	*Staphylococcus aureus*	No[64]
KKH-Sa_Cas9	N_2NRRT (R = A or G)	Engineered from Sa_Cas9.	No[64]
Bl_Cas9	N_4CND (D = A, G or T)	*Brevibacillus laterosporus*	No[64]
CasX	TTCN	*Deltaproteobacteria*	Yes[65]
CasY	NTA	*Katanobacteria*	Yes[65]
As_Cpf1	TTTV	*Acidaminococcus sp.* BV3L6	Yes[66]
Lb_Cpf1	TTTV	*Lachnospiraceae bacterium* ND206	Yes[66]
MAD7	YTTN	*Eubacterium rectale*	Yes[67]
C2c1	TTN	*Alicyclobacillus acidoterrestris*	Yes[44]

Cpf1 CRISPR nucleases, also known as Cas12a nucleases, are members of the type V group of CRISPR systems. These are similar to Cas9 in that they are a single protein, as opposed to a large complex to perform DNA interference. Cpf1 nucleases however differ in that they naturally utilise a single gRNA molecule to associate to the nuclease, and target the complex to the DNA (Figure 1.2)[45,68]. Once the target has been identified, Cpf1 nuclease induces a staggered cut, rather than a blunt cut as with Cas9 (Figure 1.2). Furthermore, this 5-8 nt cut is introduced at the PAM distal end of the protospacer region[69]. Cpf1 nucleases cleave at T-rich PAM

sites (Table 1.1) and have the capacity to cleave both DNA and RNA. This RNA cleavage capacity allows native Cpf1 systems to mature their own pre-crRNA array transcripts following the formation of an RNA pseudoknot[45]. Similar to Cas9, Cpf1 nucleases relax and interrogate DNA following binding of gRNA, initially identifying PAM sites, and thereafter analysing the protospacer region. The critical seed region of the protospacer within Cpf1s is 5-6 nt proximal to the PAM.

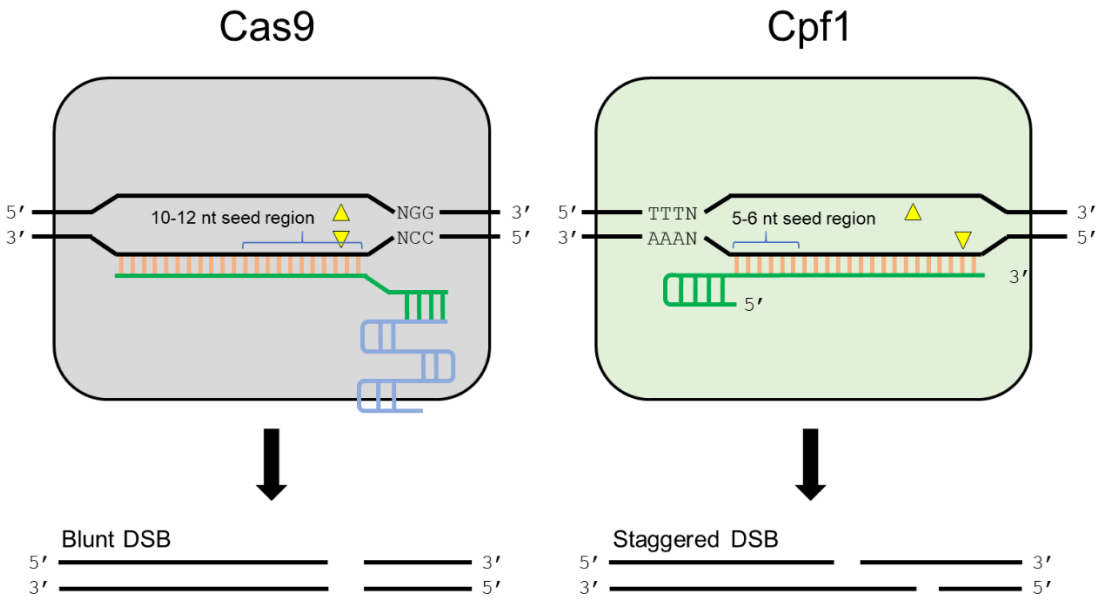


Figure 1.2 – Comparison of the basic characteristic found with native Cas9 and Cpf1 nucleases.

Cas9 requires a G-rich protospacer adjacent motif (PAM) while Cpf1 requires a T-rich PAM. Native Cas9 systems use two RNA molecules to guide the nuclease to the target location (crRNA (green) and tracrRNA (blue)), while Cpf1 uses a single guiding crRNA molecule with a shorter total length. Cas9 causes a blunt double-strand break (DSB), proximal to the PAM, while Cpf1 causes a 5 nt staggered cut, distal to the PAM (yellow triangles indicate cleavage sites). The seed region required for Cas9 binding and cleavage is indicated for both sets of nucleases.

Recently, Inscripta Inc. (USA) released an alternative CRISPR nuclease isolated from *Eubacterium rectale* with the trade name, MAD7. This nuclease exhibits many of the same characteristics as Cpf1 nucleases, with *Acidaminococcus sp.* BV3L6 (AsCpf1) being the closest known relative at the time of its release (31%

identity)[67,70]. Indeed, MAD7 has since been identified as having a closer homology to other Cpf1 nucleases, and has been designated by some as ErCas12a[71].

MAD7 has the potential to be of significant commercial interest with Inscripta claiming it is free to use in commercial R&D as well as there being no read-through royalties on any strains developed using MAD7, so long as the final strain no longer contains a copy of the nuclease or any derivatives[67]. This is in stark contrast to Cas9 and other Cpf1 nucleases which are surrounded by relative confusion in terms of their use within commercial enterprises and who provides a licence for their use. This, alongside potentially costly read-through royalties on strains developed with Cas9 or Cpf1 nucleases could have limited the level of commercial exploitation of CRISPR genome engineering within small-medium sized commercial enterprises, and in large-scale fermentations where profit margins can be slim depending on the product.

1.3.4 Transcriptional regulation with catalytically inactive variants of Cas9

The engineering of dCas9 has led to the development of two new branches of research: CRISPR interference (CRISPRi), and CRISPR activation (CRISPRa)[61,72]. CRISPRi refers to inhibition of gene transcription by associating the dCas9 to the promoter or 5´ end of a gene, which, in bacteria, sterically hinders the RNA polymerase, preventing transcription. This can be a powerful tool as studies have shown the effect of CRISPRi can be comparable to having a full knock out (KO) of the gene in question in both eukaryotes and prokaryotes[61,72]. CRISPRi has also been used as a knockdown tool to lower the transcription of essential genes which are otherwise impossible to KO without supplementing the host with either the product the deleted protein creates or adding either an exact copy or heterologous homolog of the deleted gene[72,73].

CRISPRi has been further exemplified within Cpf1 based systems following inactivation of the DNA cleaving function within the DNase dead variants (ddCpf1 E993A)[68]. Within *E. coli* it was found that targeting the template strand within a target gene yielded greater transcriptional repression efficiencies (~330-fold repression for three gRNA) than when the non-template strand was targeted

(~6-fold for two gRNA and no repression for a third) when compared to a control strain without a targeting gRNA[74].

CRISPRa on the other hand stimulates stronger transcription of a gene[59,75,76]. Here, dCas9 is fused to a transcription effector molecule and targeted to a region upstream of the promoter elements. This stimulates the recruitment of the RNAP through interactions with the fused effector, for example VPR in *Saccharomyces cerevisiae*[77], the ω RNAP subunit in *Escherichia coli* mutants in which this subunit is not expressed[59], and the ω or α RNAP subunits within *B. subtilis*[60]. When using dCas9 for CRISPRi or dCas9 fused to an effector protein for CRISPRa, and multiple simultaneously transcribed sgRNAs, the knockdown or up regulation of several genes (multiplexing) can be achieved in parallel[59–61,72,75,78].

In *B. subtilis* these have recently been combined to repress transcription of genes encoding extracellular proteases (for a similar outcome to strains described in section 1.2.2.3), while increasing the transcription of the primary extracellular chaperon encoding gene, *prsA*. dCas9-ω was targeted upstream of *prsA* (upregulation) and within the coding sequence for *bpr*, *vpr* and *nprB* (downregulation). This system, combined with an oligonucleotide annealing based promoter shuffling (OAPS) technique, was used to increase the heterologous extracellular production level of *B. licheniformis* α-Amylase (BLA) by 260-fold (100-fold by OAPS and a further 2.6-fold by CRISPRi/CRISPRa)[60].

1.3.5 CRISPR-enabled deaminase base editing for genomic editing without a double-strand break or donor DNA

The fusion of effector proteins to dCas9 is not limited to CRISPRi or CRISPRa. Deaminase subunits, such as the activation-induced cytidine deaminase (AID) enzyme from humans (encoded by *AICDA*)[79] or the evolved *E. coli* adenine deaminase, TadA[80], have also been fused to dCas9 or nCas9 to enable editing without the need for DSB repair. CRISPR-enabled base editing facilitates nucleobase deamination by hydrolysing the amino group from deoxycytosine (C) or deoxyadenosine (A), resulting in the formation of deoxyuridine (U) or inosine (I) respectively. These are subsequently altered to thymidine (T) and deoxyguanosine (G) respectively during DNA replication, provided the deaminated base is not removed before DNA replication can utilise it as a template[81,82]. To date, CRISPR

enabled deaminase base editing in prokaryotes has been exemplified in *E. coli*[80,83], *Brucella melitensis*[83], *Corynebacterium glutamicum*[54,84], and in this study (Chapter 6) within *B. subtilis*.

1.4 Mechanisms enabling CRISPR-Cas editing

Cas nucleases cleave DNA causing a DSB. This must be repaired by NHEJ or by HDR, when dDNA is present, for the host to survive[47]. Alternatively, HR of the dDNA can occur prior to the DSB. Since dDNA encodes a programmed target site edit, together with a mutation disrupting the PAM, Cas nucleases are no longer able to cut the hosts DNA at the target location, allowing the host to survive so long as there are no additional, off-target, DSB events. Thus, in the later scenario, Cas nucleases act as a counterselection mechanism against cells without the recombined genotype, as opposed to the initiating cause requiring HDR or NHEJ to proceed.

While eukaryotes have efficient NHEJ and HDR repair mechanisms, in most bacteria and yeast HDR is the only viable repair pathway. In *E.* coli, this is due to the lack of the proteins responsible for NHEJ, the DNA-binding protein Ku and the ligase/polymerase LigD[75,81]. In *Saccharomyces cerevisiae* and some prokaryotes, such as *B. subtilis*, NHEJ is only active at certain points in the cell cycle[81,85,86]. In *B. subtilis*, cell-cycle dependent NHEJ is active when the cells carry one single copy of the chromosome (1C) for long periods of time – late in stationary phase, or when sporulating. In exponential growth phase, due to constant DNA replication and the desire to produce a viable spore with 1C, the cell ensures it has >1C, mediated by the proteins Sda (preventing sporulation while DNA replication is active) and SirA (prevents new rounds of DNA replication when the cell enters sporulation)[81]. However, efficient HR systems, such as that present in *B. subtilis* could also lead to the prevention of DSB formation due to dDNA recombining and disrupting the DNA cleavage site, prior to CRISPR-Cas interrogation. This, along with the potential benefit to HR efficiency from the CRISPR-Cas system, is investigated within this study (section 3.3.3).

1.4.1 Prevention is better than cure: homologous recombination can prevent lethal double-strand breaks following CRISPR-Cas interrogation

In most bacteria, including *B. subtilis*, a DSB is lethal and thus the repair process, or the avoidance of the cut being introduced is vital[81]. The ability of the CRISPR-Cas system to continually reintroduce the DSB following accurate repair of the DNA acts as a counterselection mechanism. Thus, HR of dDNA carrying the mutation of interest as well as a synonymous PAM site mutation is vital for the cell to survive.

As described in section 1.2.1, double-stranded dDNA used within *B. subtilis* transformation experiments is incorporated to the cell in a ssDNA manner following nuclease degradation (Figure 1.3)[22]. ssDNA is protected within the cell by DprA, SsbA, SsbB and RecN[22]. The ATPase recombinase, RecA, is added to the 5´ of the ssDNA by RecN, forming a nucleoprotein filament. RecA scans the chromosome for homologous DNA and causes strand invasion. This is followed by branch migration, enabled by RecG or RuvAB, and subsequent formation of Holliday junctions (Figure 1.3)[87]. Holliday junctions are resolved following cleavage by RecU. The replaced recipient ssDNA strand is degraded within the cell and the resulting strain contains a heteroduplex with different alleles on each strand (Figure 1.3)[87]. Following DNA replication and daughter cell formation, one cell will contain the recombined genotype, while the other will contain the original parental genotype. If the introduced change provides a dominant phenotype, the subsequent population will be enriched for the modified genotype.

Following such an event, CRISPR-Cas systems confer a counterselection against the daughter cells with the parental genotype, thus ensuring only modified cells survive to form colonies. However, in *B. subtilis* there has yet to be experimental evidence presented to show if HR or HDR is the primary mechanism for CRISPR-Cas enabled genome editing (investigated in section 3.3.3).

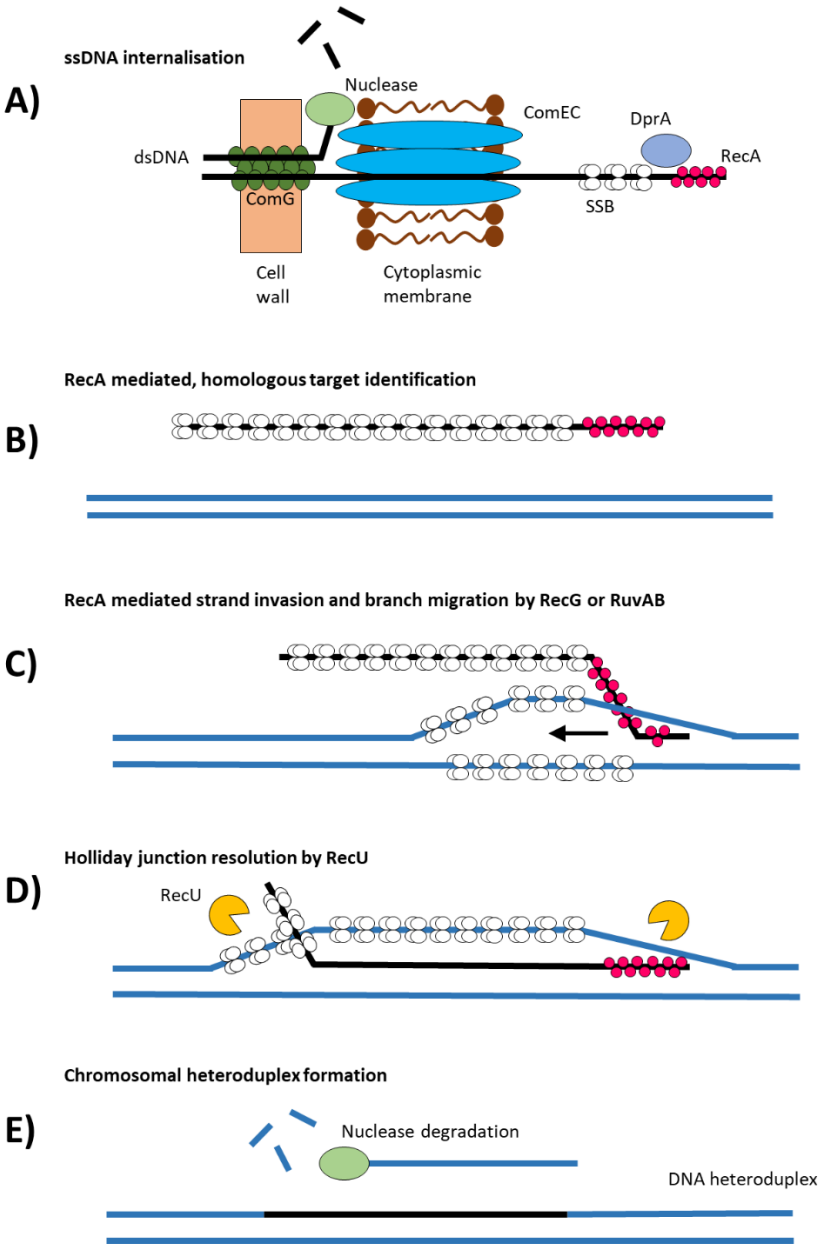


Figure 1.3 – Homologous recombination pathway following natural competency mediated transformation of *B. subtilis*.

A) dsDNA is bound by ComG pseudopilus and the ComEC channel internalises a single strand while the other is degraded. Single strand binding (SSB) proteins SsbA and SsbB bind and protect the ssDNA before DprA facilitates the addition of RecA to the 5´ of the ssDNA, replacing SSB. **B)** RecA-ssDNA proteofilament identifies target with homology to ssDNA upon the chromosome. **C)** RecA mediates strand invasion and RecG or RuvAB mediates branch migration. **D)** RecU mediates resolution of Holliday junctions. **E)** The replaced recipient strand is degraded within the cell and a DNA heteroduplex is formed. Subsequent DNA replication and cell division results in one daughter cell with the inserted sequence, and one with wild-type sequence.

1.4.2 Homology directed repair

In the absence of the CRISPR-Cas mechanism, *B. subtilis* efficiently repairs DNA DSBs, in the vegetative state, by HR. DNA damage can happen regularly by stalled DNA replication and by environmental factors, such as ionising radiation. Thus, an efficient and accurate method of DNA repair is essential to maintain genetic stability. Due to this efficiency, it is possible that DSB formation could increase the level by which dDNA recombination can take place.

DSB repair is initiated by RecN which responds early (<3 seconds[88]), and PNPase (encoded by *pnpA*) following DNA DSB (Figure 1.4)[89]. Recent single molecule imaging by Rösch et al. showed RecN diffusion within the cell changes dramatically, following DSB induction by the DNA damaging agent mitomycin C, from being very diffuse to rapidly scanning the DNA for DSBs[88]. RecN forms repair complexes at DNA DSB sites and is thought to bind the 3´-OH site of the DSB, preventing DNA degradation by non-specific exonucleases[88].

The RecN complex causes the association of the AddAB complex which in turn disassociates RecN from the DSB (Figure 1.4). AddA is a helicase and endonuclease which cleaves the 3´-5´ strand. AddB on the other hand does not have a helicase function, only acting as a nuclease for the 5´-3´ strand. AddAB degrades both strands until a Chi site is reached, which in *B. subtilis* consists of the sequence 5´-AGCGG-3´. Following the Chi site, to which AddB binds, the DNA continues to be unwound by AddAB, while only the 3´-5´ strand continues to be degraded by AddA, releasing an eventual ssDNA as a substrate for the recombinase RecA to bind. SsbA, however, initially protects this ssDNA from nucleases and prevents the formation of secondary structures[90]. This is then displaced by RecO and RecR which additionally facilitates, alongside RecF (forming the RecFOR complex), the binding of the RecA recombinase, to the ssDNA forming the RecA-ssDNA nucleoprotein filament (Figure 1.4)[88,90,91].

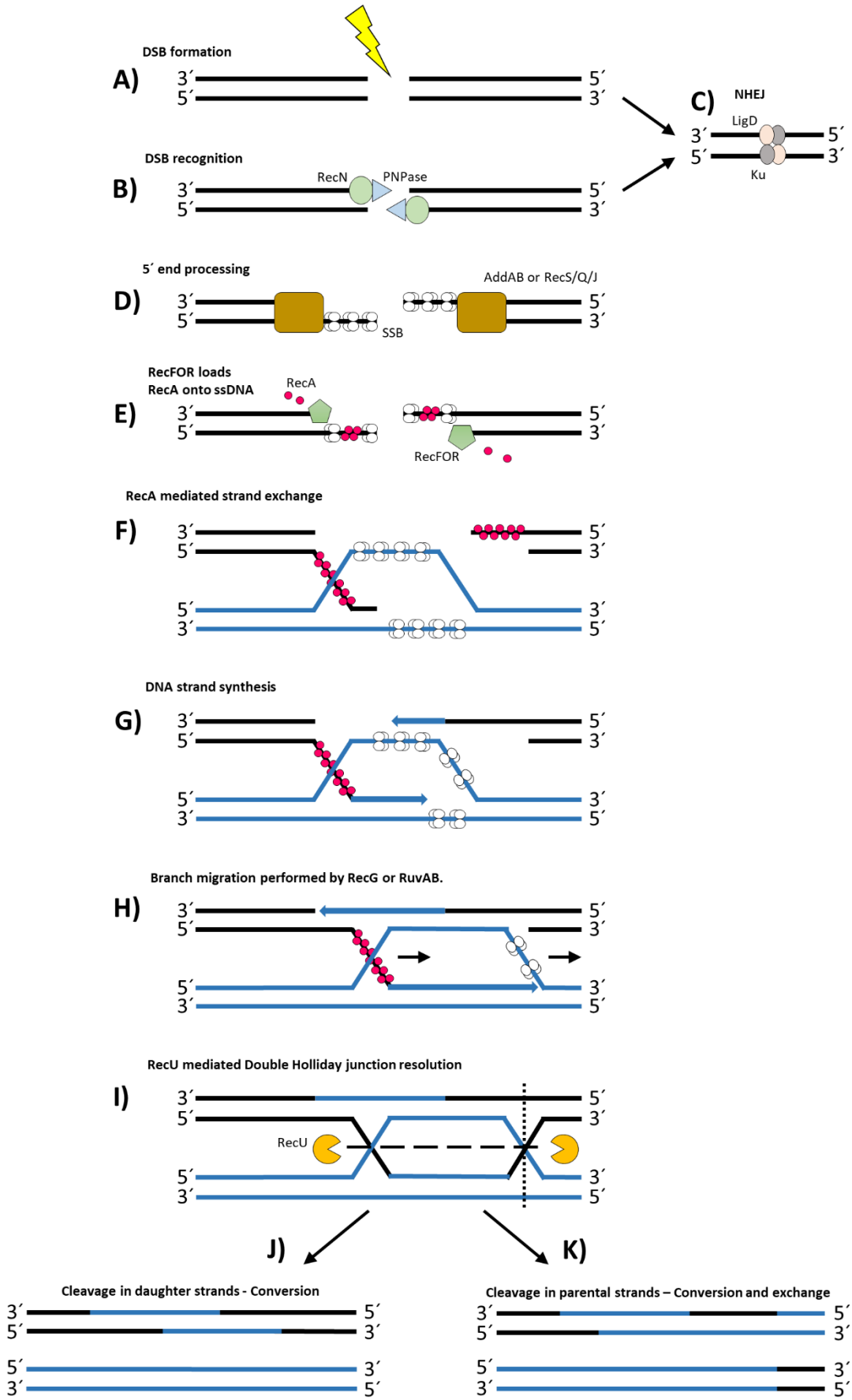

DSB formation
A)
3´
5´
5´
3´
C)
NHEJ
LigD
Ku
DSB recognition
B)
RecN
PNPase
5´ end processing
D)
AddAB or RecS/Q/J
SSB
RecFOR loads
RecA onto ssDNA
E)
RecA
RecFOR
RecA mediated strand exchange
F)
DNA strand synthesis
G)
Branch migration performed by RecG or RuvAB.
H)
RecU mediated Double Holliday junction resolution
I)
RecU
J)
K)
Cleavage in daughter strands - Conversion
Cleavage in parental strands – Conversion and exchange

Figure 1.4 – Model for DNA double-strand break repair in *B. subtilis* (previous page).

A) Double strand break (DSB) created in the genome. **B)** DSB is identified and 'repair centres' are established by RecN and PNPase. **C)** The DSB can be repaired using the error-prone non-homologous end joining (NHEJ) system, consisting of a dsDNA binding protein (Ku) to bridge the gap and recruit a ligase (LigD) to complete the DSB repair. **D)** the ends are processed by the AddAB helicase/nuclease complex, or perhaps alternatively by RecQ or RecS helicases and RecJ nuclease. Both strands are degraded until a Chi site is reached, after which only the 5´-3´ strand is further degraded. The exposed ssDNA is bound by SSB (Single strand binding protein). **E)** RecFOR mediates the disassociation of SSB and loading of RecA, forming the ssDNA-RecA nucleoprotein filament. **F)** The nucleoprotein filament undergoes a homology search and forms a displacement loop (D-loop) upon pairing with a template, displacing one strand from the template. **G)** Using the homologous strand as a template, the 3´ ends of each strand are extended by DNA polymerase. **H)** RecG or RuvAB can cause strand migration, increasing the degree of strand exchange. **I)** Double Holliday Junctions are resolved by RecU. **J)** If RecU cleaves the daughter strand (dashed line in I), DNA conversion results where the flanking sequences are the same. **K)** If RecU cleaves the parental strand (dotted line in I), DNA conversion and exchange results where the sequence downstream of the DSB are exchanged between the two strands.

In *B. subtilis* DNA replication initiation occurs rapidly during growth in nutrient rich media, termed nutrient-mediated growth rate regulation[92], ensuring ample dDNA for HR following DSB events. RecA catalyses the hydrolysis of ATP which, in the presence of SsbA and SsbB stimulates strand exchange in regions of complementarity. Branch migration is then performed by RecG or RuvAB while DNA polymerase extends the 3´ end to fill the gaps in the DNA strands[93], using the homologous strand as a template (Figure 1.4). Once polymerisation is complete, the strands are sealed by a ligase, forming a double Holliday junction. In the resolution of these Holliday junctions by RecU, equal numbers of conversion or conversion and exchange products are expected, depending on the strand which is cleaved by RecU (Figure 1.4)[94,95].

1.4.3 Non-homologous end joining

The expression of the Ku and LigD proteins, responsible for NHEJ, is regulated, in *B. subtilis*, by the forespore-specific sigma factor G (σ^G)[96]. σ^G initiates expression of 81 genes, including the DNA binding protein SpoVT. SpoVT initiates

the expression of 20 genes, including Ku and LigD[96,97]. Under sporulation conditions, such as extreme temperatures or lack of nutrients, σ^G is expressed and is responsible for late forespore gene expression[98]. σ^G acts as an auto regulator, in combination with σ^F (early forespore gene expression) enabling further expression during the spore life-span and allowing efficient genome maintenance[96,98].

Crucial to NHEJ is the binding of the Ku homodimer prior to basal DSB end-processing, by the proteins RecN and PNPase – the first steps in HDR[99]. NHEJ can still be utilised after DNA tethering by RecN, but once DSB ends have been processed to expose a 3´ tail and RecA binds, the cell is committed to HDR[99,100]. Ku binds and bridges the dsDNA ends while recruiting the DNA ligase LigD which joins the two DNA strands (Figure 1.4)[99,100]. *B. subtilis* Ku was confirmed to be functional in DSB repair *in vivo*, by KO mutants for Ku and RecA shown to be very susceptible to DSB forming agents[101]. Additionally, Wang *et al*. showed that Ku is compartmentalised to the forespore during its expression. It was confirmed *in vitro* that Ku enables an increase in the polymerisation and ligation efficiency of LigD of gaps ≥ 2 nt[101].

NHEJ has, however, been reported to not play a role within CRISRP-Cas9 editing within *B. subtilis*[102]. Furthermore, Zheng *et al*. investigated the ability of *B. subtilis* Ku and LigD for DNA repair in *E. coli*, and found them not to enable repair of the Cas9 induced DSB. This was in contrast to *Mycobacterium smegmatis* Ku and LigD which were found to be highly efficient for DSB repair, enabling large genomic deletions (67 kbp and 123 kbp)[103].

1.4.4 Base excision repair

In bacteria, the base excision repair (BER) pathway fixes non-bulky lesions in DNA. BER is thought to be the most commonly used DNA repair mechanism due to many sources having the capacity to introduce these lesions – alkylation, oxidation, depurination/depyrimidation, deamination and dUTP incorporation during DNA replication[81]. Lesions are detected and the damaged base removed by glycosylases hydrolysing the N-glycosidic bond. This results in the formation of an abasic (AP) site which is in turn recognised by AP endonucleases and AP lyases, breaking the phosphodiester bond at the 5´ and 3´ sides of the AP site respectively. The AP site is subsequently processed by an exonuclease or a

deoxyribophosphodiesterase. The small gap is closed by DNA polymerase and ligated to restore the damaged base to its undamaged form[81].

1.5 Utilisation of CRISPR systems within Bacilli

The well understood and efficient mechanisms and tools for natural competence and HR in *B. subtilis* may account for the relatively slow uptake of the CRISPR-Cas9 genome editing toolkit, in comparison to other hosts with less efficient HR, such as *Escherichia coli*[43] or *Lactobacillus reteri*[104]. However, the first example of the use of CRISPR-Cas9 in *B. subtilis* came in 2016 by Jakutyte-Giraitiene and Gasiunas. Here *Streptococcus thermophilus cas9*, and one of two sgRNA expression cassette targeting the bacteriophage SPP1 were genomically integrated into the *amyE* locus of *B. subtilis* YB886. A third nonsense sgRNA strain was included as a control. These guides allowed a reduction in sensitivity to SPP1 bacteriophage by 5-6 orders of magnitude, proving the ability for Cas9 to target non-native DNA in *B. subtilis*[105]. This is commercially significant due to the susceptibility of industrial scale fermentations to bacteriophage contamination[105].

Westbrook *et al*. demonstrated CRISPR-Cas9 use in *B. subtilis* 168 using genomic integration of Cas9 in the *lacA* locus. The sgRNA expression cassette was integrated at the 'recombination hotspot' *thrC* with the glucose repressed *mazF* counter-selectable marker. In both cases, an antibiotic selection marker was used to select for genomic integration. Utilising the toxic nature of *mazF*, the sgRNA was flanked by short direct repeats which allowed a single crossover event to occur in the absence of glucose, thus removing the sgRNA expression cassette and *mazF gene.* A new sgRNA expression cassette could then be introduced at the same locus[106]. The same study found that a chromosomally maintained and expressed Cas9 does not affect hyaluronic acid production[106]. Multiplexing was analysed and a high efficiency of single KO for two genes, *ugtP* and *amyE* (85% and 86% respectively), and double KO for both was achieved (85%) using optimised HDR conditions: a plasmid based homology template for *amyE*, PCR based homology template for *ugtP*; 1000 bp homology arms; and PAM site close to gene start codon[106]. Lastly, the authors developed a CRISPRi method targeting repression of *lacZ* expression. Using RT qRT-PCR they found 8 fold less *lacZ* mRNA and a corresponding 8 fold loss of β-galactosidase activity[106]. While this was a comprehensive toolbox for CRISPR-Cas9 activity in *B. subtilis*, representing the first

such example for gene KO or KI (knock-in), the need for genomic integration limits the speed with which experiments can be carried out. Westbrook *et al.* went on to utilise this genome editing approach to enhance heterologous production of hyaluronic acid by the streptococcal hyaluronan synthase in *B. subtilis*. CRISPRi was utilised to downregulate the expression of both *pfkA* and *zwf* to improve the titre of hyaluronic acid by 108%[107]. They subsequently went on to engineering the cell membrane, utilising dCas9 to downregulate the level of *ftsZ*, known to initiate cell division, which caused a dispersion in the cardiolipin within the membrane since its synthase, ClsA, localises to FtsZ[108].

Similarly, Peters *et al*. developed a specific and titratable CRISPRi protocol for gene knock down in *B. subtilis*. *dcas9* was integrated into the *lacA* locus, while gRNAs were integrated at *amyE* or *thrC*. All 289 known (258) and proposed (31) essential genes within *B. subtilis* were analysed. It was established that dCas9 inhibition acts upon genes downstream of the target gene within an operon, upstream genes were also affected. Growth curves and cell morphology microscopy allowed the determination of gene knock-down effect *in vivo*, providing a framework for essential gene functional analysis in Gram-positive organisms[73]. CRISPRa has also been exemplified within *B. subtilis* recently by Lu *et al*., as discussed in section 1.3.4[60].

Altenbuchner, developed a plasmid based system used for gene KO and KI with Cas9[63]. Using a single plasmid system containing *S. pyogenes* Cas9, sgRNA expression cassette and dDNA for HDR with 700 bp homology arms, 4.1 kbp and 25.1 kbp deletions were introduced to *B. subtilis* 168. Additionally, introduction of an isoleucine codon in the *trpC2* gene cured the tryptophan auxotrophic nature of the strain. Once all editing was completed, the plasmid was removed with a heat sensitive origin of replication, pE194[63]. This tool was then utilised by Watzlawick and Altenbuchner to stably incorporate up to five copies of the β-galactosidase encoding gene, *ganA*, to the chromosome of *B. subtilis* and showing this to be more stable than a strain containing a multicopy plasmid expressing *ganA*[109]. Toymentseva and Altenbuchner also developed a set of CRISPR plasmids for use in other Bacilli, based on the heat sensitive pE194 origin. These enabled the use of CRISPR-Cas systems by conjugation, which is generally utilised in strains of *B. subtilis* or other Bacilli where natural competency is inefficient or totally absent[102]. Subsequently, the CRISPR-Cas9 plasmid initially developed by Altenbuchner was utilised to alter bacteriophage DNA *in vivo* to enable viral phenotypic investigations[110]. Recently,

Yi *et al*. utilised the Altenbuchner method for investigation of *B. subtilis* HS3, and adapted the plasmid for use with *Bacillus mycoides* EC18 to investigated the interactions of these rhizosphere-associated bacteria with plant roots[111].

Zhang *at al.* developed an alternative plasmid based Cas9/sgRNA with dDNA to improve the industrial applicability of an industrial relevant, poorly transformable strain, *B. subtilis* ATCC 6051a by lowering the level of foam (*srfC*), removing extracellular proteases (*nprE* and *aprE*) and α-amylase (*amyE*), as well as the ability to form spores (*spoIIAC*), by deleting the genes responsible for each[112]. The final strain produced 2.5-fold more of the industrially relevant enzyme β-cyclodextrin glycosyltransferase.

Other CRISPR-Cas9 systems have been developed for large genomic deletions. Burby and Simmons developed a single plasmid system for *cas9* and sgRNA expression, as well as containing the dDNA (similar to the method developed by Altenbuchner) to delete the 30.5 kbp PBSX prophage which is induced following DNA damage and has the ability to cause cell lysis[113]. So *et al*., developed a two-plasmid system in which *B. subtilis* was transformed with the *cas9* expressing plasmid, before making this strain competent and transforming with a second plasmid expressing a sgRNA and carrying dDNA. Using this approach, a 38 kbp plipastatin-synthesising *pps* operon was deleted (80% efficient), as well as *spo0A* deletion (100% efficient), point mutations (68% efficient) and GFP insertion in *sigE* (97% efficient)[114].

Recently, Li *et al*. developed a method for genome editing in *B. licheniformis*. Their system incorporated nCas9 within the chromosome, before electroporating the subsequent strain with replicative plasmids containing dDNA and a sgRNA expression cassette. A 42.7 kbp region was deleted (79% efficiency), two genes were simultaneously deleted (11.6% efficiency), the *aprN* gene from *B. subtilis* was inserted (76.5% efficient) and lastly extracellular protease genes (*wprA*, *mpr*, *aprE*, *vpr*, *epr*) were sequentially deleted[115]. Liu *et al*. created a similar system for use within *B. subtilis* recently. A two plasmid system, one expressing nCas9 while the second expressed sgRNA(s) and harboured dDNA, enabled efficiencies of 80% for DNA deletion (1-8 kbp), 90% for 1-2 kbp insertions, and up to 65% for three simultaneous point mutations when *ligD* was deleted from the cell[116].

Lastly, Mougiakos *et al*. developed a CRISPR-Cas9 method for use within the facultative thermophile, *Bacillus smithii*. This host is industrially attractive since

its ability to grow in elevated temperatures reduces cooling costs, increases substrate and product solubility, reduces contamination risk, and is the optimal temperature for enzymatic lignocellulosic degradation enabling simultaneous saccharification and fermentation[117]. This study utilised a single plasmid to express *cas9* and sgRNA, while also containing dDNA. The high natural recombination efficiency was used to incorporate the dDNA while Cas9 was heat-inactivated at 45-55 °C. The temperature was subsequently lowered to 37 °C and Cas9 was used to counterselect cells which had not undergone HR of the dDNA[117].

1.6 Aims

The study presented here aims to improve upon the tools currently available by combining the flexibility of plasmid based CRISPR systems and the use of rapidly constructed dDNA in the form of a PCR product. This increases the rate at which experiments can be implemented as it removes the requirement for cloning of the dDNA within the plasmid expressing the Cas nuclease and gRNA. This co-transformational approach is made possible by the high rate of natural competence inducible within *B. subtilis*. This can, for example, allow the rapid *in situ* engineering of industrially relevant proteins, such as subtilisin E, with a single targeting plasmid and a library of dDNA. As it had not been established in *B. subtilis* if HDR increased the editing efficiency in *B. subtilis*, or if HR, prior to the DSB, alone prevented DNA cleavage, this was also investigated.

The exploitation of the industrially significant CRISPR nuclease, MAD7, within *B. subtilis* was investigated and engineered to enable licence and royalty free methods of CRISPR enabled genome editing and CRISPRi. This has the capacity to overcome current licencing restrictions associated to Cas9 and Cpf1 nucleases which can be a significant disincentive to its use for industrial fermentation strain construction. These tools were utilised to initiate the construction a heterologous protein expression library through sequential knock-out of extracellular proteases.

Chapter 2 – Materials and methods

In this chapter, the experimental techniques and materials used throughout this study are described.

2.1 Bacterial strains and plasmids

All bacterial strains and plasmids used in this study are described at the start of the relating chapter. Schematic representations of the major plasmids used in this study can be found in Appendix A. All cloning of plasmids was carried out in *Escherichia coli* Top10 (Invitrogen (*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu) 7697 galU galK rpsL* (Str^R) *endA1 nupG*)).

2.2 Growth and maintenance of bacterial cultures

2.2.1 General growth and storage of cultures

Lysogeny broth (LB) medium was used for all general-purpose *E. coli* and *Bacillus subtilis* growth cultures. No salt LB medium was used for *E. coli* electrocompetent cell preparation (Section 2.4.1). *B. subtilis* was transformed by the induction of natural competence following the method described by Bennallack *et al*. (Section 2.4.2). Growth media components are described fully in Appendix B**.** Growth was monitored by measuring optical density at 600 nm (OD_{600}) in a spectrophotometer. When growth was desired on a solid agar plate, select agar was added to 1.5% (w/v) during medium preparation.

Unless otherwise stated, growth cultures were performed in non-baffled 125 mL Erlenmeyer flasks at 37 °C with orbital shaking at 250 rpm. When appropriate, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mmol L^{-1} to induce protein expression. When appropriate, antibiotics were added to the relevant concentration as described in Table 2.1.

Unless otherwise stated, to ensure comparability between strains and replicates, growth was synchronised through the use of a pre-culture to grow diluted overnight cultures to exponential growth phase, in the same conditions as the assay

culture. Generally, overnight cultures were diluted to OD_{600nm} 0.05 for subsequent pre-culture growth.

For long term storage of bacterial cultures, overnight LB cultures were supplemented to 10% (v/v) glycerol. These were stored at -80 °C in cryovials.

Table 2.1 – Stock and working solutions of antibiotics.

Antibiotic	Solvent	Stock concentration ($mg.mL^{-1}$)	Working concentration ($\mu g.mL^{-1}$)
Ampicillin (Amp)	H_2O	200	200
Chloramphenicol (Cm)	100% (v/v) ethanol	34	10
Spectinomycin (Spec)	H_2O	100	200
Kanamycin (Kan)	H_2O	100	10

2.2.2 Platereader growth curves and GFP detection

Due to the autofluorescence associated with LB medium an alternative medium with lower autofluorescence was used for CRISPRa experiments where small changes in GFP expression could be encountered. CHG medium was used when small changes in GFP expression levels were being detected[118]. Growth media components are described fully in Appendix B**.**

2 mL LB medium overnight cultures, supplemented with the appropriate antibiotics, in 24-well plates, grown at 37 °C, 250 rpm, were used to inoculate 200 µL CHG medium pre-cultures to OD600 0.05. These were subsequently grown to mid-exponential phase (OD_{600} 0.6-0.8) at 37 °C, 300 rpm. At this point, when the cell states have been normalised to increase the reproducibility of experiments, the pre-cultures were used to inoculate the pre-warmed assay cultures. Growth curves and GFP detection was carried out in 96-well U-bottom plates (Greiner Bio-One, Austria) using the FLUOstar Omega Microplate Reader (BMG LABTECH, Germany). 200 µL cultures were incubated at 37 °C, 300 rpm. OD_{600} was measured

every 10 minutes. GFP was detected every 10 minutes at 485/520 nm (excitation/emission) and gain was set at 1264.

2.2.3 RoboLector analysis

To perform growth curves and detect small changes in GFP expression levels, the RoboLector XL microfermentation system (m2p-labs, Germany) was used. Biomass and GFP were monitored.

Prior to RoboLector XL analysis, single colonies were used to inoculate 2 mL CHG medium cultures, supplemented with the appropriate antibiotics, and grown overnight in 24-well plates. These were used to inoculate 2 mL pre-warmed CHG medium in 24-well plates to OD_{600} 0.1 and subsequently were grown to mid-exponential phase (OD_{600} 0.6-0.8) at 37 °C, 200 rpm in a Multitron Standard Infors HT (Switzerland) incubator. The pre-cultures were diluted to OD_{600} 0.1 in 1 mL pre-warmed CHG medium in 48-well FlowerPlate®. These plates were incubated in the RoboLector XL at 37 °C, 800 rpm, 85% humidity. Biomass (excitation: 620 nm, gain: 4) and GFP (excitation: 488 nm, emission: 520 nm, gain: 10) were monitored with measurements taken every 15 minutes.

2.2.4 Minimal media for tryptophan auxotroph curing selection

To select colonies for prototrophic phenotype, M9 agar plates were prepared with the appropriate antibiotics and glucose as the carbon source. 20 mg/L Tryptophan was also included where appropriate as a control to ensure all other components of the medium were included. Growth media components are described fully in Appendix B. Prior to spreading transformation colonies (Section 2.4.2), the cells were washed three times with phosphate buffered saline (10 mmol L^{-1} phosphate buffer, 2.7 mmol L^{-1} potassium chloride, 137 mmol L^{-1} sodium chloride, pH 7.4).

2.2.5 Cell Growth Quantifier system for growth analysis

The effect of TadA-nCas9 on growth rate, when compared to TadA-dCas9, was analysed using the Cell Growth Quantifier (CGQ) system (Aquila Biolabs GmBH, Germany). The CGQ system measures light backscatter of growth cultures

to determine culture density. As the LEDs and photodiodes for measurements are situated *in situ* underneath the shake flask, there is no need to remove the cultures from incubation. Measurements were taken every 60 seconds.

Growth rate (μ) was determined by the CGQ software (A.U./min) using Equation 2-1 in which two time points (t_1 and t_2) are compared and X is the cell density (determined by light backscatter) at each time point (X_{t1} and X_{t2}).

Equation 2-1 – Growth rate

$$\frac{ln\frac{X_{t2}}{X_{t1}}}{(t_2 - t_1)}$$

Since the small differences between shake-flasks can affect the backscatter reading, growth was zero'd to allow comparison between replicates by the subtraction of the first measurement from each subsequent reading.

2.3 Chemicals

All chemicals and reagents were obtained from Merck (USA) unless otherwise stated.

2.4 Bacterial transformations

2.4.1 Preparation and transformation of electrocompetent *Escherichia coli*

E. coli, as well as other species, can be transformed with DNA by the addition of an electrical charge to promote migration of DNA across the cell membrane. To prepare electrocompetent *E. coli* cells, a 400 mL No Salt LB media culture, in a 2 L baffled Erlenmeyer flask was inoculated to OD_{600} 0.1 from an overnight LB culture and grown at 37 °C, 250 rpm to mid exponential phase (OD_{600} 0.5 – 0.7). The cells were then chilled on ice before harvesting by centrifugation and undergoing two washes with ice cold sterile H_2O. The cells were washed once more with ice cold, sterile 20% (v/v) glycerol. The cells underwent a final resuspension in

4 mL ice cold, sterile 20% (v/v) glycerol before being aliquoted in ice cold 1.5 mL tubes and storing at -80 °C until required for use.

To transform the electrocompetent *E. coli*, 50 µL of competent cells were defrosted on ice and mixed with up to 5 µL of the ice-cold DNA containing solution. This mixture was added to a 1 mm electroporation cuvette. A 1.7 kV, 200 Ω and 25 µF electric current was applied to the cuvette using a Bio-Rad (USA) GenePulser electroporator. 945 µL S.O.C. medium was added to the cuvette and the cell suspension was transferred to a 1.5 mL tube. This was incubated at 37 °C, 250 rpm for 1 hour before the cells were spread on an LB agar plate with the appropriate antibiotics and incubated overnight at 37 °C.

2.4.2 Preparation and transformation of competent *Bacillus subtilis*

The natural competence pathway well described in *B. subtilis* based on the expression of the master competency regulator protein, ComK, was utilised to prepare competent *B. subtilis*. The method described by Bennallack *et al*. was followed, with some modifications[24]. Information on growth medium components can be found in Appendix B.

A single colony was used to inoculate 10 mL LB medium in a 125 mL non-baffled flask, with the appropriate antibiotics where required, and incubated overnight at 37 °C, 250 rpm. In a 125 mL flask, 14 mL starvation medium 1 (SM1) was inoculated with 1 mL of the overnight culture and grown until the culture enters stationary phase. 15 mL pre-warmed starvation medium 2 (SM2) was added and grown for a further 90 minutes under the same conditions. At this point, and for the next 2 hours, the cells are highly competent. In a 15 mL tube, 500 µL of cells were mixed with the DNA to transform the cell and incubated at 37 °C, 250 rpm for 30 minutes. 300 µL of LB medium was added and incubated further at 37 °C, 250 rpm for 30 minutes. The transformation mixture was then spread on an LB agar plate with the appropriate antibiotics and incubated overnight at 37 °C.

2.5 DNA manipulation

General molecular biology techniques for DNA manipulation were performed as per Sambrook and Russell (2001)[119]. More detail is provided in the following sections.

2.5.1 Purification of plasmid DNA

Plasmid DNA was purified from *E. coli* overnight cultures using the Monarch® Plasmid Miniprep Kit (New England BioLabs® Inc. (NEB), USA) following the manufacturer's instructions. This kit uses a column-based system that contains a silica membrane to which plasmid DNA can be bound, washed and eluted. All plasmid DNA was eluted with molecular biology grade H_2O. 5 mL of bacterial culture was spun down and resuspended in 200 µL Plasmid Resuspension Buffer. 200 µL Plasmid Lysis Buffer was added and mixed well by inversion before incubating for 1 minute. 400 µL Plasmid Neutralisation Buffer was added and mixed well by inversion before incubation at room temperature for 2 minutes. This lysate was clarified by centrifugation at 16,000 rcf for 5 minutes and the supernatant was added to a spin column. The plasmid DNA was bound to the silica by centrifugation at 16,000 rcf for 1 minute. The column was washed with 200 µL Plasmid Wash Buffer 1 by centrifugation at 16,000 rcf for 1 minute, before being washed again with 400 µL of Plasmid Wash Buffer 2 by centrifugation at 16,000 rcf for 1 minute. The DNA was eluted with 50 µL of DNA Elution Buffer.

When a large volume of plasmid DNA was required, the QIAGEN® (Germany) Plasmid Midi Kit was utilised as per the manufacturer's instructions. 100 mL of bacterial culture was spun down and resuspended in 6 mL Buffer P1. 6 mL Buffer P2 was added to this, mixed by inversion and incubated at room temperature for 5 minutes. 6 mL chilled Buffer P3 was added to this, mixed by inversion and poured into QIAfilter Cartridge, with a cap screwed into the tip, before 10 minutes incubation at room temperature. A HiSpeed Midi Tip was equilibrated by adding 4 mL Buffer QBT, which was allowed to pass through the column under gravity. The tip was removed from the QIAfilter Cartridge the cell lysate was filtered into the equilibrated HiSpeed Tip by gently pushing a plunger into the QIAfilter Cartridge. The lysate was allowed to enter the resin of the HiSpeed Tip under gravity. The HiSpeed Tip was then washed with 20 mL Buffer QC, before eluting the

DNA with 5 mL Buffer QF. The DNA was precipitated by adding 3.5 mL isopropanol to the eluted DNA, mixed and incubated at room temperature for 5 minutes. This mix was then gently passed through a QIAprecipitator using a syringe using a constant pressure. Using the same syringe, the DNA bound to the QIAprecipitator was washed with 2 mL 70% v/v ethanol. Air was then passed through the QIAprecipitator several times to ensure the ethanol was fully removed. Using a fresh syringe, the DNA was eluted in 1 mL TE Buffer and passed though the QIAprecipitator several times to ensure maximum DNA recovery.

2.5.2 DNA quantification

DNA concentration was quantified using the Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, USA) in conjunction with the Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). The manufacturer's instructions were followed with the exception of a fraction of the DNA sample being diluted 10x prior to addition to the assay buffer to maximise the accuracy of quantification. The Qubit® dsDNA HS working solution was prepared by mixing 1 µL of the Qubit® dsDNA HS Reagent with 199 µL Qubit® dsDNA HS Buffer, per sample to be analysed as well as for two additional standards. 199 µL of this was added to each tube for analysis of samples, while 190 µL was added to each tube for standard analysis. 1 µL of the diluted sample was added to each sample tube, while 10 µL of each standard (0 ng/µL and 10 ng/µL) were added to each standard tube. Following 2 minutes incubation at room temperature and ensuring no bubbles were present in the base of the tube, these were analysed using the Qubit® 2.0 Fluorometer on the dsDNA HS setting.

2.5.3 Isolation of genomic DNA from *Bacillus subtilis*

To purify genomic DNA from *B. subtilis*, 1 mL of an LB medium overnight culture, containing the appropriate antibiotics where applicable, was harvested at 16,000 x*g* for 1 minute at room temperature. The supernatant was discarded and the pellet was resuspended in 100 µL 0.85% (w/v) sodium chloride, supplemented with 5 µL 10 mg/mL RNAse ONE (Promega, USA) and 0.8 µL 125 mg/mL Lysozyme from chicken egg white. This was incubated at 37 °C for 5 minutes. 300 µL Cell Lysate Solution (25 mmol L^{-1} ethylenediaminetetraacetic acid (EDTA), 2% (w/v) sodium dodecyl sulphate (SDS)) was added and mixed well before addition of

168 µL Protein Precipitation Solution (10 mol L^{-1} ammonium acetate). The mix underwent 20 seconds of vortexing before 5 minutes incubation on ice. The cellular debris was removed by centrifugation at 16,000 x*g* for 10 minutes before transferring the supernatant to a new tube. 600 µL 100% (v/v) isopropanol was added and mixed well. DNA was isolated by centrifugation at 16,000 x*g*. The supernatant was removed and the DNA was washed with 200 µL 70% (v/v) ethanol. Following centrifugation at 16,000 x*g* for 2 minutes, the supernatant was discarded. Any remaining ethanol was removed by incubation at 55 °C for 10 minutes. The isolated genomic DNA (gDNA) was resuspended in 100 µL molecular biology grade H_2O and stored at 4 °C.

2.5.4 Oligonucleotides

Unless otherwise stated, all single strand DNA (ssDNA) oligonucleotides used in this study were obtained from Merck. All oligonucleotides are described in Appendix C. If used for PCR, oligonucleotides were designed using the Clone Manager software (SciEd, USA). Oligonucleotides used for sgRNA protospacer cloning were designed as described in section 2.7.1. Upon delivery, oligonucleotides were resuspended in molecular biology grade H_2O. 100 µmol L^{-1} stocks were stored at -40 °C, and 10 µmol L^{-1} working solutions were stored -20 °C.

2.5.4.1 Oligonucleotide phosphorylation and annealing

ssDNA oligonucleotide pairs designed for ligation with an additional piece of prepared DNA were 5´ phosphorylated, using T4 Polynucleotide Kinase (NEB, USA), and annealed in a one-pot reaction. The reaction mixture and thermo-cycler protocol utilised for this procedure are described in Table 2.2.

Table 2.2 – Reaction mixture and conditions for ssDNA oligonucleotide phosphorylation and annealing.

Reaction Mixture			Reaction conditions	
Component	**Volume (µL)**	**Final Concentration**	**Lid Temperature: 105 °C**	
100 µmol L^{-1} Oligonucleotide 1	5	10 µmol L^{-1}	**Temperature (°C)**	**Time (minutes)**
100 µmol L^{-1} Oligonucleotide 2	5	10 µmol L^{-1}	37	30
10x T4 polynucleotide kinase buffer	5	1x	65	20
10 mmol L^{-1} ATP	5	1 mmol L^{-1}	90	5
100 mmol L^{-1} DTT	2.5	5 mmol L^{-1}	85	1
50% (w/v) PEG8000	5	5% (w/v)	80	1
10 units/µL T4 Polynucleotide Kinase	1	10 units	75	1
H_2O	21.5		70	1
			65	1
			60	0.5
			55	0.5
			50	0.5
			45	0.5
			40	0.5
			35	0.5
			30	0.5
			25	0.5
			20	3

2.5.5 Polymerase chain reaction (PCR)

DNA was amplified *in vitro* using the polymerase chain reaction (PCR).

2.5.5.1 High fidelity DNA amplification

DNA to be used in plasmid cloning or genomic intergradation was prepared using Phusion® High-Fidelity DNA polymerase (NEB) as per the manufacturer's instructions (Table 2.3). The thermo-cycling conditions for a routine PCR reaction using Phusion® High-Fidelity DNA polymerase is described in Table 2.4. Primer annealing temperatures were determined using the NEB T_m calculator online tool.

Table 2.3 – Standard PCR reaction using Phusion® High-Fidelity DNA polymerase.

Component	Volume (µL)	Final concentration
5x Phusion HF buffer	10	1x
10 mmol L^{-1} dNTPs	1	200 µmol L^{-1}
10 µmol L^{-1} forward primer	2.5	0.5 µmol L^{-1}
10 µmol L^{-1} reverse primer	2.5	0.5 µmol L^{-1}
Template DNA	Variable	< 250 ng
2 units/µL Phusion® High-Fidelity DNA polymerase	0.5	1 unit
H_2O	Up to 50 µL	

Table 2.4 – Standard thermo-cycling reaction conditions for use with Phusion® High-Fidelity DNA polymerase.

Step	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 seconds	1
Denaturation	98	10 seconds	35 cycles
Primer annealing	45 – 72	20 seconds	
Extension	72	15 seconds per kbp	
Final extension	72	5 minutes	1

2.5.5.2 Colony PCR (cPCR)

PCR was also used to identify colonies were cloning of a DNA fragment into a vector had been successful, or to identify where colonies had successfully had a region of DNA deleted or inserted. Here, a picked colony is suspended in 10 µL LB medium with the relevant antibiotic and used as the PCR template. This colony PCR (cPCR) procedure was performed using One*Taq*® Quick-Load® 2x Master Mix with Standard Buffer (NEB, USA). See Table 2.5 for a routine reaction mix composition, and Table 2.6 for routine thermo-cycling reaction conditions when using One*Taq*® Quick-Load® 2x Master Mix with Standard Buffer. The remaining colony suspension from any successful constructs was used to streak the strain out on LB agar with the appropriate antibiotic.

Table 2.5 – Standard cPCR reaction using OneTaq® Quick-Load® 2x Master Mix with Standard Buffer.

Component	Volume (µL)	Final concentration
One*Taq*® Quick-Load® 2x Master Mix with Standard Buffer	10	1x
10 µmol L^{-1} forward primer	1.5	0.75 µmol L^{-1}
10 µmol L^{-1} reverse primer	1.5	0.75 µmol L^{-1}
Colony suspension template DNA	Variable	
H_2O	Up to 20 µL	

Table 2.6 – Standard cPCR thermo-cycling reaction conditions for use with OneTaq® Quick-Load® 2x Master Mix with Standard Buffer.

Step	Temperature (°C)	Time	Cycles
Initial denaturation	94	5 minutes	1
Denaturation	94	30 seconds	30 cycles
Primer annealing	45 – 68	30 seconds	
Extension	68	1 minute per kbp	
Final extension	68	5 minutes	1

2.5.5.3 Overlap extension PCR

Two DNA fragments, prepared using PCR, are fused in a two-step process, known as overlap extension PCR (OE-PCR). These two DNA fragments share 20 bp with 100% homology at the extreme end where the two fragments are to be fused[120]. The two DNA fragments are added at the same molar ratio. This region of homology (~20 nt) acts as priming regions during the first PCR phase (Table 2.7 and Table 2.8). During a pause in the thermo-cycling program, oligonucleotide primers are added hybridising to the 5´ end of the first fragment, and the 3´ end of the second fragment. During the second PCR phase, these primers amplify the final full construct. This product is isolated for subsequent purification by agarose gel electrophoresis (Section 2.5.6.1).

Table 2.7 – Standard OE-PCR reaction using Phusion® High-Fidelity DNA polymerase.

	Component	Volume (µL)	Final concentration
Phase 1	5x Phusion HF buffer	10	1x
	10 mmol L^{-1} dNTPs	1	200 µmol L^{-1}
	Template PCR-1	Variable	100 ng
	Template PCR-2	Variable	
	2 units/µL Phusion® High-Fidelity DNA polymerase	0.5	1 unit
	H_2O	Up to 45 µL	
Phase 2	10 µmol L^{-1} forward primer	2.5	0.5 µmol L^{-1}
	10 µmol L^{-1} reverse primer	2.5	0.5 µmol L^{-1}

Table 2.8 – Standard OE-PCR thermo-cycling reaction conditions for use with Phusion® High-Fidelity DNA polymerase.

	Step	Temperature (°C)	Time	Cycles
Phase 1	Initial denaturation	98	30 seconds	1
	Denaturation	98	10 seconds	15 cycles
	Primer annealing	45 – 72	40 seconds	
	Extension	72	15 seconds per kbp	
Pause to add primer oligonucleotides				
Phase 2	Denaturation	98	10 seconds	20 cycles
	Primer annealing	45 – 72	40 seconds	
	Extension	72	15 seconds per kbp	
	Final extension	72	5 minutes	1

2.5.6 Electrophoresis of DNA

DNA was analysed by agarose gel electrophoresis, or capillary gel electrophoresis.

2.5.6.1 Agarose gel electrophoresis

DNA grade agarose (Severn Biotech Ltd., United Kingdom) was suspended in 100 mL Tris-Acetate-EDTA (TAE) buffer (40 mmol L^{-1} Tris, 20 mmol L^{-1} Acetic acid, 1 mmol L^{-1} EDTA) and dissolved by microwave heating. 7 µL SYBR® Safe DNA Gel Stain (Thermo Fisher Scientific, USA) was added to enable visualisation of the DNA. The gel was cast with the required well size determined by the volume of the DNA containing solution following addition of gel loading dye, purple (NEB, USA) to 1x concentration. Once set, the DNA containing solution was added and the loaded gel underwent electrophoresis at 100 V for 30 – 45 minutes. DNA bound with SYBR® Safe DNA Gel Stain was visualised with the Safe Imager™ 2.0 Blue Light Transilluminator (Thermo Fisher Scientific, USA) system (excitation: 470 nm, emission: 530 nm). DNA band sizes were determined by comparison with 2-Log DNA Ladder (0.1-10.0 kbp) (NEB, USA).

2.5.6.2 Capillary gel electrophoresis

Analytical DNA analysis, were DNA did not need to be recovered, was performed using the Agilent 2200 TapeStation System (Agilent Technologies, USA). DNA bands are separated on a capillary agarose column, increasing the speed with which samples can be analysed. Furthermore, the sensitivity of DNA detection is increased to 5 pg/µL if required. DNA sizes up to 1 kbp were analysed with the D1000 Screentape and electronic DNA ladder. DNA sizes up to 5 kbp were analysed with the D5000 Screentape and comparison with the D5000 DNA marker. As per the manufacturer's instructions, 1 µL of the DNA containing sample was mixed with 3 µL (D1000) or 10 µL (D5000) of the appropriate buffer. Results were analysed using the 2200 TapeStation Controller Software (Agilent Technologies, USA), and Agilent TapeStation Analysis software (Agilent Technologies, USA). The buffers contain a low bp DNA marker and a high bp DNA marker with which the analysis software can directly compare between different capillaries.

2.5.7 Purification of DNA from agarose gel

DNA isolated by agarose gel electrophoresis (Section 2.5.6.1) was excised using a scalpel. Excess agarose gel not harbouring DNA was removed and the DNA was subsequently purified using the Monarch® DNA Gel Extraction Kit (NEB, USA). This kit utilises the standard bind/wash/elute workflow with a silica DNA binding matrix which lowers buffer transfer when compared to other common comparable kits. The gel slice was added to a pre-weighed 1.5 mL tube, the subsequent weight was measured and 4 volumes of Monarch Gel Dissolving Buffer was added (1 mg of gel equates to 1 µL buffer). The mix was incubated at 55 °C until the gel slice had fully dissolved, and subsequently added to a spin column. The DNA was bound to the silica by centrifugation at 16,000 rcf for 1 minute. The DNA was then washed twice with 200 µL DNA Wash Buffer, centrifuging for 1 minute at 16,000 rcf after each buffer addition. The DNA was eluted in into a fresh 1.5 mL tube with 10 µL molecular biology grade H_2O or DNA Elution Buffer.

2.5.8 Digestion of DNA with restriction nucleases

DNA fragments can be digested by restriction endonucleases in a blunt or staggered manner to expose blunt or 'sticky ends'. Type IIP restriction endonucleases cleave the DNA within their targeting region, leaving a consistent region to facilitate cloning of DNA fragments. Type IIS restriction endonucleases on the other hand cleave the DNA outside their recognition site. Thus the 'sticky end' can be different between different DNA fragments cut with the same type IIS restriction endonucleases. Ligases allow the joining of DNA fragments at a compatible site. For example, bluntly cleaved DNA can be ligated to any other bluntly cleaved DNA fragment, provided a free phosphate group is available to facilitate the joining of the DNA backbone. Compatible 'sticky ends' are required for ligases to combine DNA fragments cut in such a manner[121]. DNA to be digested with restriction endonucleases were carried out following the manufacturer's recommendations. Unless otherwise stated, all restriction nucleases were obtained from NEB.

Parental CRISPR plasmid backbone preparations during gRNA protospacer cloning underwent restriction digestion reactions using AarI (Thermo Fisher Scientific, USA) or SapI. The AarI reaction mixture outlined in Table 2.9 was prepared and incubated for 5 hours at 37 °C, before the enzyme was heat inactivated at 65 °C for 20 minutes.

Table 2.9 – Typical reaction mixture for the preparation of parental CRISPR plasmid backbones, based on pBAC0008, during gRNA protospacer cloning.

Component	Volume (µL)	Final concentration
Plasmid DNA	Variable	Up to 1 µg
10x Buffer AarI	2	1x
50x AarI Oligonucleotide	0.4	1x
2 units/µL AarI	0.5	1 unit
H_2O	Up to 20 µL	

2.5.9 Dephosphorylation of DNA

Backbone DNA for use in cloning was 5´ dephosphorylated using Alkaline Phosphatase, Calf Intestinal (CIP) (NEB, USA) to prevent self-ligation with compatible ends. The manufacturer's protocol was followed to set up 20 µL reactions as per Table 2.10. This was incubated at 37 °C for 30 minutes before the DNA was isolated and purified by agarose gel electrophoresis.

Table 2.10 – DNA dephosphorylation reaction using Alkaline Phosphatase, Calf Intestinal (CIP).

Component	Volume (µL)	Final concentration
Linear DNA	Variable	Up to 1 pmol
Cutsmart Buffer (10x)	2	1x
1 units/µL CIP	1	1 unit
H_2O	Up to 20 µL	

2.5.10 Ligation of DNA fragments

DNA fragments were ligated using T4 DNA Ligase (NEB, USA). The reaction mixture, with a total DNA amount of 100 ng, is described in Table 2.11. Reactions were incubated at room temperature for 2 hours, heat inactivation of the ligase at 65 °C for 10 minutes and used to transform *E. coli* cells (Section 2.4.1).

Table 2.11 – Ligation reaction conditions using T4 DNA Ligase.

Component	Volume (µL)	Final concentration
Backbone DNA	Variable	Variable (mol L^{-1})
Insert DNA	Variable	5x mol L^{-1} of backbone DNA
10x T4 DNA Ligase Buffer	2	1x
400 units/µL T4 DNA Ligase	1	400 units
H_2O	Up to 20 µL	

2.5.11 Gibson DNA Assembly

The Gibson Assembly technology was used to assemble multiple DNA fragments with overlapping regions in a one-pot isothermal reaction[122]. Typically, 20 bp of homology between adjacent fragments were designed. NEBuilder® HiFi DNA Assembly Master Mix (NEB, USA) was used for all assembly reactions. The reaction master mix was diluted in an equal volume of the DNA fragments to be assembled. Insert DNA was combined with the vector in a 3:1 molar ratio. All reactions were incubated at 50 ºC for 60 min before being used to transform *E. coli* cells (Section 2.4.1).

2.5.12 InABLE® DNA Assembly

The InABLE® DNA assembly method was used to construct the initial parental plasmids to be subsequently used alongside traditional cloning methods for final plasmid construction and use within CRISPR experiments (Section 2.5)[123]. Individual 5´ truncated parts were prepared by PCR and subjected to restriction digestion at 5´ and 3´ regions with SapI. These parts were ligated to phosphorylated and annealed oligonucleotides at each terminus, containing 3 nt and 16 nt single strands at the 5´ and 3´ ends, respectively. The part- oligonucleotides fusions were annealed at the homologous 16 nt overhangs for 1 hour at 37 °C, and used to transform electrocompetent *E. coli* (Section 2.4.1).

2.5.13 Ligation Independent cloning

Ligation independent cloning (LIC) was used for the introduction of P_{liaG} for *gfpmut3* expression on the plasmid pGFPbglS, yielding plasmid pBAC0091[124]. The method outlined by Bisicchia *et al*. was followed[124]. pGFPbglS contains a LIC site: TTTTACCGCGGGCTTT**CCCGGG**AAGGAGGAACT. pGFPbglS is linearised with SmaI (recognition sequence in bold) and 4 pmol is treated with 20 units of T4 polymerase (NEB) in the presence of 1x NEBbuffer 2.1 and 2.5 mmol L^{-1} dATP (Thermo Fisher Scientific) for 20 minutes at 22 °C followed by 20 minutes at 75 °C to inactivate the enzyme. This generates single-stranded overhangs on either side of the SmaI site extending to the underlined A bases.

A P_{liaG} containing fragment was prepared by PCR using oligonucleotides oMAP0298/0299, introducing a 5´-CCGCGGGCTTTCCCAGC-3´ tail sequence via oMAP0298 and a 5´-GTTCCTCCTTCCCACC-3´ tail sequence via oMAP0299. 0.2 pmol PCR fragments were then treated with 0.4 units of T4 DNA polymerase in the presence of 1x NEBuffer 2.1 and 2.5 mmol L^{-1} dTTP at 22 °C for 20 minutes, and the enzyme was inactivated at 75 °C for 30 minutes. This generates an insert fragment with complementary single-stranded regions to those generated in pGFPbglS.

5 ng of the treated pGFPbglS was mixed with 15 ng of the treated PCR product and incubated at room temperature for 10 minutes before transformation of *E. coli* by electroporation (section 2.4.1).

2.6 DNA sequencing

DNA Sanger sequencing was primarily carried out by Source Bioscience (Cambridge, United Kingdom). Sequencing for CRISPR enabled DNA deamination base editing was performed either by Source Bioscience, GeneWiz (Beijing, China) or General Biosystems (Beijing, China).

2.7 CRISPR-Cas9 mediated gene editing in *Bacillus subtilis*

2.7.1 sgRNA protospacer design

The protospacer regions were initially identified using the online tool CRISPR-ERA: a comprehensive designer tool for CRISPR genome editing, (gene) repression, and activation, and selected based on the proximity to the desired modification[125]. This tool is only applicable for Cas9 based CRISPR experiments.

Latterly, a widely used online tool, CHOPCHOP, was adapted for use with *B. subtilis* and was used for protospacer design[126,127]. This tool is applicable for both Cas9 and AsCpf1 protospacer identification. Due to the similarities between the PAM sites for AsCpf1 and MAD7, CHOPCHOP can also identify appropriate sites for MAD7 based experiments, however gRNA scores relevant for AsCpf1 may not be relevant for MAD7 until further characterised.

2.7.2 sgRNA selection

When a gene or locus input is loaded to the protospacer identification software, all candidates are scored and listed based on% GC, predicted efficiency, the detection of any off-target effects, and the level of self-complementarity to identify any stem-loops in the subsequent RNA. Additionally, the position where the protospacer is situated is experiment dependent.

2.7.2.1 CRISPR enabled gene editing and DNA deaminase base editing

When a region was to be deleted, a PAM site was selected equidistance from the 5´ and 3´ ends of the relevant area. PAM sites used for point mutations or DNA insertions were selected for their proximity to the intended editing/insertion site.

2.7.2.2 CRISPRi

PAM sites were selected for their proximity to the gene promoter or 5´ region of the target gene.

2.7.2.3 CRISPRa

Protospacers were selected at regions extending upstream of the start codon, as was exemplified in *E. coli* by Bikard *et al*.[59]. Protospacers within the promoter act as controls for the binding of the Cas9 fusion protein to show the knock-down of the target gene.

2.7.3 Cloning of protospacer region into parental plasmids

Parental plasmids were digested with either AarI or SapI, depending on the parental plasmid used (Section 2.5.8), before being dephosphorylated to prevent self-ligation (Section 2.5.9). Two DNA oligonucleotides were annealed and phosphorylated (Section 2.5.4.1). The annealed oligonucleotides consisted of the protospacer region for targeting of the Cas proteins and 4 nt or 3 nt on the 5´ and 3´ ends to generate a single stranded overhang for compatibility with the AarI or SapI, respectively, digested parental plasmid. These were subsequently ligated (Section 2.5.10) with the prepared parental plasmid. Insertion was confirmed by cPCR

(Section 2.5.5.2) and sequencing with oMAP0148. All final plasmids used for CRISPR approaches are listed in Table 2.12.

Table 2.12 – Final CRISPR plasmids used throughout this study.

Plasmid ID	Parental plasmid	CRISPR associated protein	Target gene	PAM			Oligonucleotides used for cloning protospacer DNA
				Sequence (5´-3´)	Distance from		
					Start codon	-35 promoter element	
pBAC0027	pBAC0015	Cas9	*amyE*	CGG	147 bp downstream	-	oMAP0089/0091
pBAC0035	pBAC0015	Cas9	NT	-	-	-	oMAP0145/0147
pBAC0041	pBAC0015	Cas9	*amyE*	AGG	25 bp downstream	-	oMAP0125/0127
pBAC0047	pBAC0015	Cas9	*amyE*	CGG	83 bp downstream	-	oMAP0140/0142
pBAC0054	pBAC0015	Cas9	*aprE*	CGG	359 bp downstream	-	oMAP0150/0151
pBAC0055	pBAC0015	Cas9	*aprE*	AGG	1163 bp downstream	-	oMAP0156/0157

pBAC0065	pBAC0015	Cas9	*rpoZ*	AGG	143 bp downstream	-	oMAP0208/0209
pBAC0085	pBAC0015	Cas9	*trpC2*	AGG	347 bp downstream	-	oMAP0694/0695
pBAC0092	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	CGG	-	113 bp upstream (Pveg and PrrnB P1)	oMAP0224/0225
pBAC0093	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	AGG	-	131 bp upstream (Pveg and PrrnB P1)	oMAP0226/0227
pBAC0094	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	AGG	-	139 bp upstream (Pveg and PrrnB P1)	oMAP0228/0229
pBAC0095	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	TGG	-	163 bp upstream (Pveg and PrrnB P1)	oMAP0230/0231
pBAC0096	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	AGG	-	200 bp upstream (Pveg and PrrnB P1)	oMAP0232/0233
pBAC0097	pBAC0067	dCas9-ω(Es)	*gfpmut3*	AGG	-	163 bp upstream (PliaG)	oMAP0324/0325
pBAC0098	pBAC0067	dCas9-ω(Es)	*gfpmut3*	GGG	-	125 bp upstream (PliaG)	oMAP0326/0327

pBAC0099	pBAC0067	dCas9-ω(Es)	*gfpmut3*	AGG	-	145 bp upstream (PliaG)	oMAP0328/0329
pBAC0100	pBAC0067	dCas9-ω(Es)	*gfpmut3*	TGG	-	154 bp upstream (PliaG)	oMAP0330/0331
pBAC0101	pBAC0067	dCas9-ω(Es)	NT	-	-	-	oMAP0145/0147
pBAC0102	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	GGG	-	80 bp upstream (PliaG) 16 bp upstream (Pveg and PrrnB P1)	oMAP0320/0321
pBAC0103	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	AGG	-	97 bp upstream (PliaG) 33 bp upstream (Pveg and PrrnB P1)	oMAP0322/0323
pBAC0104	pBAC0082	dCas9-ω(Bs)	NT	-	-	-	oMAP0145/0147
pBAC0110	pBAC0090	dCas9-Φ29	*gfpmut3*	GGG	-	80 bp upstream (PliaG)	oMAP0320/0321
pBAC0111	pBAC0090	dCas9-Φ29	*gfpmut3*	AGG	-	97 bp upstream (PliaG)	oMAP0322/0323
pBAC0112	pBAC0090	dCas9-Φ29	*gfpmut3*	GGG	-	125 bp upstream (PliaG)	oMAP0326/0327
pBAC0113	pBAC0090	dCas9-Φ29	*gfpmut3*	AGG	-	145 bp upstream (PliaG)	oMAP0328/0329
pBAC0114	pBAC0090	dCas9-Φ29	*gfpmut3*	TGG	-	154 bp upstream (PliaG)	oMAP0330/0331

pBAC0115	pBAC0090	dCas9-Φ29	NT	-	-	-	oMAP0145/0147
pBAC0120	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	CGG	-	84 bp upstream (PliaG)	oMAP0336/0337
pBAC0121	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	GGG	-	73 bp upstream (PliaG)	oMAP0338/0339
pBAC0122	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	TGG	-	68 bp upstream (PliaG)	oMAP0340/0341
pBAC0123	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	TGG	-	56 bp upstream (PliaG)	oMAP0342/0343
pBAC0124	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	CGG	-	45 bp upstream (PliaG)	oMAP0344/0345
pBAC0125	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	CGG	-	30 bp upstream (PliaG)	oMAP0346/0347
pBAC0126	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	CGG	-	23 bp upstream (PliaG)	oMAP0350/0351
pBAC0127	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	CGG	-	11 bp upstream (PliaG)	oMAP0354/0355
pBAC0129	pBAC0015	Cas9	*gfpmut3*	GGG	18 bp upstream	-	oMAP0386/0387
pBAC0132	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	CGG	-	20 bp upstream (Pveg)	oMAP0406/0407
pBAC0133	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	GGG	-	61 bp upstream	oMAP0326/0327

						(Pveg and PrrnB P1)	
pBAC0134	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	AGG	-	20 bp downstream (PrrnB P1)	oMAP0412/0413
pBAC0135	pBAC0082	dCas9-ω(Bs)	*gfpmut3*	GGG	-	9 bp upstream (PrrnB P1)	oMAP0414/0415
pBAC0160	pBAC0159	AsCpf1	NT	-	-	-	oMAP0553/0554
pBAC0161	pBAC0159	AsCpf1	*amyE*	TTTG	21 bp downstream	-	oMAP0547/0548
pBAC0162	pBAC0158	MAD7	*amyE*	TTTG	21 bp downstream	-	oMAP0549/0550
pBAC0163	pBAC0158	MAD7	NT	-	-	-	oMAP0555/0556
pBAC0165	pBAC0015	Cas9	*gfpmut3*	TGG	27 bp downstream	-	oMAP0573/0574
pBAC0166	pBAC0158	MAD7	*gfpmut3*	TTTC	21 bp downstream	-	oMAP0571/0572

pBAC0167	pBAC0159	AsCpf1	*gfpmut3*	TTTC	21 bp downstream	-	oMAP0569/0570
pBAC0168	pBAC0164	dCas9	*amyE*	TGG	47 bp downstream	-	oMAP0563/0564 and oMAP0565/0566
			gfpmut3	GGG	68 bp upstream	-	
			NT	-	-	-	
pBAC0169	pBAC0041	TadA-dCas9	*amyE*	AGG	25 bp downstream	-	oMAP0125/0127
pBAC0170	pBAC0041	TadA-nCas9	*amyE*	AGG	25 bp downstream	-	oMAP0125/0127
pBAC0171	pBAC0041	dCas9-AID	*amyE*	AGG	25 bp downstream	-	oMAP0125/0127
pBAC0172	pBAC0041	nCas9-AID	*amyE*	AGG	25 bp downstream	-	oMAP0125/0127
pBAC0184	pBAC0041	dCas9	*amyE*	AGG	25 bp downstream	-	oMAP0125/0127

pBAC0185	pBAC0015	Cas9	*trpC2*	CGG	368 bp downstream	-	oMAP0694/0695
pBAC0188	pBAC0162	dMAD7	*amyE*	TTTG	21 bp downstream	-	oMAP0549/0550
pBAC0189	pBAC0163	dMAD7	NT	-	-	-	oMAP0555/0556
pBAC0190	pBAC0166	dMAD7	*gfpmut3*	TTTC	21 bp downstream	-	oMAP0571/0572
pBAC0192	pBAC0175	dCas9-AID	*nicK*	GGG	40 bp downstream	-	oMAP0682/0683
pBAC0193	pBAC0176	nCas9-AID	*nicK*	GGG	40 bp downstream	-	oMAP0682/0683
pBAC0194	pBAC0035	dCas9	NT	-	-	-	oMAP0145/0147
pBAC0195	pBAC0165	dCas9	*gfpmut3*	TGG	27 bp downstream	-	oMAP0573/0574
pBAC0197	pBAC0158	MAD7	*aprE*	TTTG	12 bp downstream	-	oMAP0701/0702

pBAC0198	pBAC0158	MAD7	*nprE*	TTTA	6 bp downstream	-	oMAP0705/0706
pBAC0199	pBAC0158	MAD7	*htrA*	TTTC	22 bp downstream	-	oMAP0709/0710
pBAC0200	pBAC0158	MAD7	*htrB*	TTTG	25 bp downstream	-	oMAP0713/0714
pBAC0201	pBAC0158	MAD7	*bpr*	TTTT	1 bp upstream	-	oMAP0724/0725
pBAC0202	pBAC0158	MAD7	*nprB*	TTTT	2 bp upstream	-	oMAP0731/0732
pBAC0203	pBAC0158	MAD7	*mpr*	TTTG	1 bp upstream	-	oMAP0738/0739
pBAC0204	pBAC0158	MAD7	*epr*	TTTC	6 bp downstream	-	oMAP0745/0746
pBAC0205	pBAC0158	MAD7	*wprA*	TTTT	15 bp downstream	-	oMAP0752/0753
pBAC0206	pBAC0158	MAD7	*vpr*	TTTT	7 bp downstream	-	oMAP0717/0718

pBAC0207	pBAC0187	dMAD7	*gfpmut3*	TTTA	9 bp downstream	-	oMAP0759/0760
pBAC0208	pBAC0187	dMAD7	*gfpmut3*	CTTC	16 bp downstream	-	oMAP0761/0762
pBAC0209	pBAC0187	dMAD7	*gfpmut3*	CTTG	43 bp downstream	-	oMAP0763/0764
pBAC0210	pBAC0187	dMAD7	*gfpmut3*	TTTG	78 bp downstream	-	oMAP0765/0766
pBAC0211	pBAC0187	dMAD7	*gfpmut3*	TTTC	80 bp downstream	-	oMAP0767/0768
pBAC0212	pBAC0187	dMAD7	*amyE*	TTTG	4 bp downstream	-	oMAP0769/0770
pBAC0213	pBAC0187	dMAD7	*amyE*	TTTG	11 bp downstream	-	oMAP0771/0772
pBAC0214	pBAC0187	dMAD7	*amyE*	TTTA	27 bp downstream	-	oMAP0775/0776

pBAC0215	pBAC0187	dMAD7	*amyE*	TTTA	51 bp downstream	-	oMAP0777/0778
pBAC0218	pBAC0158	MAD7	*trpC2*	TTTC	299 bp downstream	-	oMAP0799/0800
pBAC0219	pBAC0187	dMAD7	*gfpmut3*	TTTC	80 bp downstream	-	oMAP0793/0794
			gfpmut3	TTTC	21 bp downstream	-	
pBAC0220	pBAC0187	dMAD7	*gfpmut3*	TTTC	80 bp downstream	-	oMAP0797/0798
			amyE	TTTG	4 bp downstream	-	
pBAC0222	pBAC0187	dMAD7	*amyE*	TTTG	4 bp downstream	-	oMAP0795/0796
			amyE	TTTA	51 bp downstream	-	

pBAC0223	pBAC0018	dCas9	*nicK*	GGG	40 bp downstream	-	oMAP0682/0683
pBAC0224	pBAC0041	nCas9	*amyE*	AGG	25 bp downstream	-	oMAP0125/0127

2.7.4 Donor DNA (dDNA) preparation

Linear dDNA was prepared by OE-PCR (Section 2.5.5.3), where the modifications of interest were introduced into the primers used for the initial PCR products. In the event that the modifications of interest have been introduced into a strain during a previous experiment, the gDNA for this strain (following curing of the editing plasmid) was purified, and used for PCR amplification.

2.7.5 Co-transformation of *B. subtilis*

Naturally competent *B. subtilis* (Section 2.4.2), were transformed with 200 ng of the editing plasmid alongside 1 µg of dDNA at ~2 kbp in length.

2.7.6 Curing of editing plasmid from *B. subtilis*

The CRISPR editing plasmids were removed from the edited strain by promoting plasmid loss in LB supplemented with 1 mmol L^{-1} IPTG overnight. Here the presence of IPTG encourages curing by introducing the pressure of protein expression without the presence of the appropriate antibiotic to retain the plasmid within the cell.

Plasmid loss was confirmed by counter plating on LB agar plates with and without chloramphenicol before a subsequent round of editing, or other uses for the strain. Counter plating was performed by first patching a sterilised velvet cloth from an LB agar plate harbouring the colonies to be screened. This was then used to inoculate first an additional LB agar plate, and an LB agar plate with the appropriate antibiotics. Following growth at 37 °C, colonies which grew on LB agar without antibiotics, but not on the plate containing chloramphenicol, were judged to have lost the plasmid. This was confirmed by growing the colony in LB media only, and LB media with chloramphenicol.

2.8 Protein analysis by Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Proteins were visualised using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

2.8.1 Protein sample preparation

Proteins expressed intracellularly were prepared for SDS-PAGE analysis in the following manner. 1 mL cell culture of a known OD_{600} was harvested by centrifugation at 16,000 x*g* for 1 minute. The supernatant was discarded and the cell pellet was resuspended in 32.2 µL cell lysis master mix / mL culture / OD_{600}. The cell lysis master mix was made up of 30 µL BugBuster® Protein Extraction Reagent, 2 µL 30 mg/mL lysozyme from chicken egg white and 0.2 µL 2.5 units/µL Benzonase® Nuclease (Merck, USA). This was incubated at 37 °C for 20 minutes to lyse the cells before being centrifuged at 16,000 x*g* for 10 minutes to separate the soluble and insoluble intracellular protein fractions. The soluble fraction (supernatant) was separated, and the insoluble fraction was resuspended in the same volume of deionised H_2O. SDS-PAGE samples were prepared by adding 10 µL of the soluble or insoluble fraction to 12.5 µL 4x BOLT LDS Sample Buffer (Thermo Fisher Scientific, USA), 5 µL 9% (w/v) dithiothreitol, and 22.5 µL deionised H_2O before being incubated at 100 °C for 5 minutes. The sample was subsequently briefly centrifuged to ensure the full volume was in the base of the tube, before being mixed by vortexing.

Extracellular protein samples and protein purification samples were analysed directly by mixing 32.5 µL of the cell culture supernatant, or purification fraction, with the SDS-PAGE sample components as above, without the addition of H_2O. The samples were boiled, centrifuged and vortexed as above.

2.8.2 Separation of samples by SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Proteins within the prepared SDS-PAGE sample were separated based on their molecular weight using the Bolt™ Bis-Tris Plus Gel system (Thermo Fisher Scientific, USA) with 1 mm gel thickness and a gradient of 4-12% polyacrylamide concentration. Intracellular protein samples were analysed by the addition of 10 µL SDS-PAGE sample to each well. Extracellular protein samples and protein purification samples were analysed by the addition of 35 µL SDS-PAGE sample to each well. Proteins were separated by the addition of a 165 V current for 40 minutes.

2.8.3 InstantBlue™ protein visualisation

Protein samples separated using the Bolt™ Bis-Tris Plus Gel system were visualised using 20 mL of InstantBlue™ Protein Stain (Expedeon, USA). This was added directly to the gel and incubated at room temperature, with gentle agitation, for 15 minutes. Following incubation, the InstantBlue™ Protein Stain was discarded, and the gel was washed with deionised H_2O.

2.9 Enzymatic activity detection

2.9.1 Analysis of α-amylase activity in solid phase

The starch degrading α-amylase enzyme can be characterised through the use of iodine to stain non-degraded starch[27]. As such, α-amylase secretion can be detected from a colony on an LB agar plate supplemented with 1% (w/v) soluble potato starch. Colonies were first used to inoculate a fresh LB plate before adding 5 mL H_2O to the plate and washing off the colonies to allow a clear identification of a degraded starch halo. 2.5 mL of 3.57 mmol L^{-1} iodine solution was added to each plate to stain the non-degraded starch a dark blue. When the α-amylase was secreted from the colony, a clear halo can be observed.

2.9.2 Analysis of α-amylase activity in liquid phase

Using biological triplicates, 2 mL LB *B. subtilis* cultures, supplemented with the relevant antibiotics and 1 mmol L^{-1} IPTG for *dcas9* or *dmad7* expression induction, were grown for 24 hours of growth at 37 °C, 250 rpm. 100 µL samples were harvested at 2250 rcf for 5 minutes and the supernatant, containing the secreted α-amylase enzyme, was stored at -20 °C until required. In a 96-well MTP, biological triplicates were analysed by mixing 25 µL of the thawed supernatant with 100 µL of ice-cold assay solution (0.05% (w/v) soluble potato starch, 50 mmol L^{-1} tris hydrochloride, 25 mmol L^{-1} calcium chloride dihydrate, pH 6.8). This was incubated at 37 °C for 30 minutes before the reaction was stopped by the addition of 50 µL of stop solution (0.01% (w/v) iodine, 0.1% (w/v) potassium iodide, 1 mol L^{-1} tris hydrochloride). Each well was analysed at 620 nm using the FLUOstar Omega Microplate Reader (BMG LABTECH, Germany) to quantify α-amylase activity levels.

A medium blank treated in the same way was used as a base line for the reaction. Several serial dilutions were carried out to ensure the absorbance was in the linear range of the spectrophotometer.

Amylase activity is shown by a decrease in absorbance. In the presence of KI, I_2 forms I_3^- which is a soluble complex that enters the coil of the starch causing an intense blue-black colour. Therefore, when starch is degraded by amylase a lighter colour is obtained. The values obtained were inverted to make it clear where greater amylase activity was identified.

2.9.3 Analysis of protease activity

Subtilisin E variants and WT activity was determined by casein degradation as described by Cupp-Enyard[128]. Experiments were carried out with three technical replicates. The protocol was modified for use with MTP to increase the experimental rigor and throughput. 20 µL samples were added to 115 µL 0.65% (w/v) casein solution (prepared in 50 mmol L^{-1} potassium phosphate dibasic trihydrate, pH 7.5), mixed well and incubated at 37 °C for 30 minutes. 115 µL 110 mmol L^{-1} trichloroacetic acid was added and incubated at 37 °C for 10 minutes to stop the reaction. The MTP was centrifuged at 2000 x*g* for 10 minutes. 50 µL of the supernatant was added to 125 µL 500 mmol L^{-1} sodium carbonate, before the further addition of 25 µL 0.5 N Folin & Ciocalteu's phenol reagent which reacts primarily with free tyrosine released by the protease. This was mixed well and incubated at 37 °C for 30 minutes. The MTP was centrifuged at 2000 x*g* for 10 minutes before transferring 100 µL of the supernatant into a fresh MTP and recording the absorbance value at 660 nm using the FLUOstar Omega Microplate Reader (BMG LABTECH, Germany).

2.9.3.1 Residual protease enzyme activity assay

The residual protease activity was measured in the same manner as in section 2.9.3. However, prior to the addition of the enzyme fraction, the samples were incubated at 55 °C for 10, 20, 40, or 60 minutes using a thermo-cycler.

2.10 GFPmut3 detection

B. subtilis 168 was modified to express GFPmut3 which was subsequently used as a reporter protein for various methods.

2.10.1 Qualitative detection

The presence of expressed GFPmut3 within colonies or liquid culture was qualitatively detected using the Safe Imager™ 2.0 Blue Light Transilluminator and Amber Filter (Thermo Fisher Scientific, USA) system (excitation: 470 nm, emission: 530 nm).

2.10.2 Quantitative detection by plate reader

Using biological triplicates, 2 mL LB cultures, containing the relevant antibiotics and 1 mmol L^{-1} IPTG for where appropriate, were grown overnight at 37 °C, 250 rpm. In the same medium, the overnight cultures were used to inoculate 2 mL cultures in 24 well plates to OD_{600} 0.05 and grown to exponential growth phase at 37 °C, 250 rpm to normalise the cell state between cultures. These were subsequently used to inoculate a further 2 mL culture of the same medium to OD_{600} 0.05. OD_{600} levels were measured and 100 µL samples were used to measure GFPmut3 using the FLUOstar Omega Microplate Reader (BMG LABTECH, Germany) 24 hours after assay culture inoculation (GFPmut3 excitation: 485 nm, emission: 520 nm, gain: 1000). The relative expression levels of *gfpmut3* were quantitatively determined by normalisation of the GFPmut3 levels (A.U.) by OD_{600}.

2.11 *In silico* design for protein engineering of *B. subtilis* 168 subtilisin E

The crystal structures for *B. subtilis* subtilisin E (PDB ID: 1SCJ) and its *Bacillus clausii* homolog, M-protease (PDB ID: 1WSD), were overlaid using Swiss-Pdb viewer[129–131].

2.12 Subtilisin E purification by size exclusion chromatography

The three subtilisin E variants and WT protein were overexpressed in the *aprE* knock-out strain BAC0114. This strain was then transformed with plasmids pBAC0059/0060/0068/0069, resulting in strains BAC0119/0120/0121/0122 respectively. These strains were grown for 24 hours at 37 °C, 250 rpm in 20 mL LB supplemented with chloramphenicol and 1 mmol L^{-1} IPTG for *aprE* variant expression. The supernatant was clarified by centrifugation (850 rcf, 10 minutes, 4 °C) and dialysed overnight into 100 mmol L^{-1} tris hydrochloride with 150 mmol L^{-1} sodium chloride at pH 8. The dialysed supernatant was concentrated to 5 mL with Amicon Ultra-15 (10 kDa membrane (Merck, USA)), filtered through a 0.2 µm filter (Sartorius, Germany). This was loaded onto a HiPrep Sephacryl S-200 HR, 120 mL 16/60 size exclusion column (GE Healthcare, USA), eluted in the same buffer composition used for dialysis and fractions found to contain subtilisin E were pooled.

2.12.1 Protein concentration determination

Protein concentration was determined by absorbance at 280 nm. Samples were loaded into a quartz cuvette and analysed using the Cary 60 UV-Vis spectrophotometer (Agilent Technologies, USA), and the Simple Reads software (Agilent Technologies, USA). The spectrophotometer was blanked with the same buffer used during protein purification. The Beer-Lambert law (Equation 2-2) was used to determine the protein concentration using the absorbance at 280 nm, the molar extinction co-efficient (ε) of AprE (32430 for all variants), and the known pathlength (l) of 1 cm.

Equation 2-2 – Beer-Lambert Law

$$\boldsymbol{Abs_{280} = \varepsilon \times C \times I}$$

2.13 Thermal shift assay

The fluorescence-based thermal shift assay was used to determine the melting temperature (T_m) of the subtilisin E variants and wild-type[132]. The assay

reaction was prepared in a total volume of 20 µL (18 µL purified subtilisin E variant in purification buffer (section 2.11), 1 µL 100 mmol L^{-1} calcium acetate and 1 µL 1:50 SYPRO® Orange). The assay was performed with six technical replicates for each variant. The reaction was analysed using the PikoReal™ Real-Time PCR System (Thermo Fisher Scientific, USA) which recorded changes in fluorescence with increasing temperature (40 °C - 80 °C in increments of 0.2 °C, held for 6 seconds at each point).

2.14 Real-time quantitative reverse transcription PCR (RT qRT-PCR)

Real-time quantitative reverse transcription PCR (RT qRT-PCR) was used to detect differences in *gfpmut3* expression levels during CRISPRa experiments in *B. subtilis*. The analysis was performed with biological triplicates and technical duplicates using two reference genes (*gapA* and *sdhA*) to allow standardisation of the samples as per the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE)[133]. Strain BAC0297 harbouring a non-targeting CRISPRa plasmid was used as the relative quantification reference.

2.14.1 Harvesting of samples

To allow reliable comparison between replicates, samples were taken during exponential phase of the assay culture following growth synchronisation using a pre-culture. Growth was monitored with dilutions of the cell culture in 96-well U-bottom plates (Greiner Bio-One, Austria) using the FLUOstar Omega Microplate Reader (BMG LABTECH, Germany).

In 24-deep well plates, three single colonies were used to inoculate 1 mL LB medium supplemented with the appropriate antibiotic and 1 mmol L^{-1} IPTG. These were grown overnight at 37 °C, 250 rpm. The following morning, these cultures were used to inoculate pre-cultures (LB medium supplemented with the appropriate antibiotic and 1 mmol L^{-1} IPTG, 1 mL final volume) to OD_{600} 0.05 and grown until mid-exponential phase (OD_{600} 0.4 – 0.8). The pre-cultures were diluted to OD_{600} 0.05 in pre-warmed LB medium supplemented with the appropriate antibiotic and 1 mmol L^{-1} IPTG, the final volume was 1 mL. The growth of these cultures was

monitored and at mid-exponential phase 50 µL (corresponding the manufacturer's recommended biomass amount for RNA purification) of the culture was removed for immediate RNA purification.

2.14.2 RNA purification

The total RNA for each sample was prepared using the Monarch® Total RNA Miniprep Kit (NEB, USA). The purification was carried out in a biological safety cabinet, and RNAseZap® (Thermo Fisher Scientific, USA) was used to clean the biological safety cabinet, pipettes and gloves prior to the experiment to avoid contamination by RNAse enzymes. The manufacturer's protocol was followed throughout.

Cells were harvested by centrifugation (500 x*g*, 1 minute) and the supernatant was removed. Two volumes of RNA Lysis Buffer was mixed with lysozyme from chicken egg white at a final concentration of 3 mg/mL. This underwent vigorous vortexing for 10 seconds, incubated at 25 °C for 10 minutes and centrifuged for 2 minutes at 16,000 x*g*. The supernatant was transferred to a gDNA Removal Column. All subsequent centrifugation steps were performed at 16,000 x*g*.

The gDNA removal column was centrifuged for 30 seconds and to the supernatant was added an equal volume of 100% (v/v) ethanol. This was mixed by pipetting before being added to an RNA Purification Column and centrifuging for 30 seconds. The flow-through was discarded and 500 µL RNA Wash Buffer was added to the column. This was centrifuged for 30 seconds and 80 µL DNase I reaction mix was loaded to the column for an on-column DNase I treatment to remove residual DNA. Following incubation at room temperature for 15 minutes, 500 µL RNA Priming Buffer was added to the column before centrifugation for 30 seconds. 500 µL RNA Wash Buffer was loaded before centrifugation for 30 seconds. A further 500 µL RNA Wash Buffer was added before centrifugation for 2 minutes. The RNA was eluted with 50 µL Nuclease-free Water and stored at -80 °C until required.

2.14.3 RNA quantification

RNA concentration was quantified with two replicates using the Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, USA) in conjunction with the Qubit™ RNA HS

Assay Kit (Thermo Fisher Scientific, USA). The manufacturer's instructions were followed. The Qubit® RNA HS working solution was prepared by mixing 1 µL of the Qubit® RNA HS Reagent with 199 µL Qubit® RNA HS Buffer, per sample to be analysed as well as for two additional standards. 199 µL of this was added to each tube for analysis of samples, while 190 µL was added to each tube for standard analysis. 1 µL sample was added to each sample tube, while 10 µL of each standard (0 ng/µL and 10 ng/µL) were added to each standard tube. Following 2 minutes incubation at room temperature and ensuring no bubbles were present in the base of the tube, these were analysed using the Qubit® 2.0 Fluorometer on the RNA HS setting.

2.14.4 Reverse transcription

RNA was reverse transcribed to cDNA using the ProtoScript® II Reverse Transcriptase (NEB, USA). In duplicate reactions, the following was added: 6 µL purified RNA; 2 µL of the supplied Random Primer Mix at 60 µmol L^{-1}; 10 µL Protoscript® II Reaction Mix; 2 µL Protoscript® II Enzyme Mix. This mixture was incubated for 5 minutes at 25 °C, followed by 60 minutes at 42 °C. The subsequent cDNA was used for quantitative PCR reactions.

2.14.5 Real-time quantitative reverse transcription PCR (RT qRT-PCR) primer identification and validation

RT qRT-PCR reaction primers were developed to allow accurate quantification of *gfpmut3*, *gapA* and *sdhA* transcription levels. Two primer sets were tested for each target transcript. Primer set 1 for *gapA* was previously reported by Song *et al*. (2016) to be effective for RT qRT-PCR in *B. subtilis*. All other primer sets were designed using Primer3Plus using the following parameters: a product range of 100-150 bp; the primer length should be a minimum of 18 bp, a maximum of 27 bp and an optimum of 20 bp; primer melting temperature should be a minimum of 57 °C, a maximum of 58 °C and an optimum of 57.5 °C; and the GC content should be a minimum of 45%, a maximum of 55% and an optimum of 50%[134]. Primers were resuspended with sterile PCR grade H_2O in a biological safety cabinet and stored at -40 °C until required for use.

Primers were validated using gDNA isolated from strain BAC0288 (*B. subtilis* 168 *ΔbglS::aphAI Pveg gfpmut3*) as described in section 2.5.3. gDNA concentration was measured in triplicate (Section 2.5.2). RT qPCR reactions were set up as in Table 2.13. Reactions were set up using undiluted, 1 in 10, 1 in 100, 1 in 1000 and 1 in 10000 template DNA dilutions alongside a no gDNA control reaction. All reactions were performed using the PikoReal™ Real-Time PCR System (Thermo Fisher Scientific, USA), using the program outlined in Table 2.14, which allowed for the detection of an increase in fluorescence when more DNA was amplified.

Table 2.13 – RT qPCR reaction conditions used for primer validation.

Component	Volume (µL)	Final concentration
PowerUp SYBR Green Master Mix [2x] (Thermo Fisher Scientific, USA)	5	1x
10 µmol L^{-1} forward primer	0.5	0.5 µmol L^{-1}
10 µmol L^{-1} reverse primer	0.5	0.5 µmol L^{-1}
Template gDNA	2	variable
Sterile PCR grade H_2O	2	

Table 2.14 – RT qPCR thermo-cycling reaction conditions used for primer validation.

Step	Temperature (°C)	Time	Cycles
Uracil-DNA glycosylase activation	50	2 minutes	1
Dual-Lock™ DNA polymerase activation	95	2 minutes	1
Denaturation	95	15 seconds	40 cycles
Primer annealing and extension	60	1 minute	
Final Extension	60	30 seconds	1
Melting Curve	60 - 95	0.2 °C / s	1

RT qPCR reactions were analysed by melt curve analysis to validate only a single product is amplified and C_q (quantification cycle) results. C_q is determined by the point at which the fluorescence signal crosses a threshold which in turn is

determined by the PikoReal™ Real-Time PCR System as 10 times greater than that of the standard deviation of the baseline. Using the C_q results from the dilutions of the gDNA template DNA, a calibration curve was established and allowed the determination of the amplification factor, and therefore the PCR efficiency, for each primer set. To determine the amplification factor, Equation 2-3 was used. PCR primers with 100% efficiency would yield an amplification factor of 2 and would correspond to a calibration curve slope value of -3.32. The primer set for each target gene with the PCR efficiency (%) closest to 100% was selected for use in subsequent RT qRT-PCR experiments.

Equation 2-3 – PCR amplification factor

$$10^{\frac{-1}{slope}}$$

2.14.6 Real-time quantitative reverse transcription PCR (RT qRT-PCR)

RT qRT-PCR reactions were set up as outlined in Table 2.15. All reactions were performed using the PikoReal™ Real-Time PCR System (Thermo Fisher Scientific, USA), using the program outlined in Table 2.14. Two technical replicates were performed for each of three biological replicate samples. Non-template controls were analysed alongside the samples to ensure there was no contaminating DNA introduced during the assay set up. Assay gene and reference genes were analysed in parallel. Assay strains were run alongside the control strain.

Table 2.15 – RT qRT-PCR reaction conditions used for *gfpmut3* transcriptional alteration analysis.

Component	Volume (µL)	Final concentration
PowerUp SYBR Green Master Mix [2x] (Thermo Fisher Scientific)	5	1x
10 µmol L^{-1} forward primer	0.5	0.5 µmol L^{-1}
10 µmol L^{-1} reverse primer	0.5	0.5 µmol L^{-1}
cDNA	3	
sterile PCR grade H_2O	1	

2.14.7 Data processing

The RT qRT-PCR reactions each gave a C_q and T_m value. C_q values were used to compare the transcription levels of *gfpmut3*, *gapA* and *sdhA* transcription between the control strain and assay strains. T_m values are used to ensure the same PCR product is compared between reactions.

The $\Delta\Delta C_q$ method was used to determine differences in transcription levels between assay and control strains. The difference in C_q values (ΔC_q), averaged by technical replicates, between control and assay strains, as well as *gfpmut3* and *gapA* or *sdhA* was calculated as in Table 2.16.

Table 2.16 – ΔCq values derived from RT qRT-PCR reactions. BAC0297 acts as the control strain.

$\Delta C_q 1$	C_q (*gfpmut3*, assay strain) - C_q (*gapA*, assay strain)
$\Delta C_q 2$	C_q (*gfpmut3*, assay strain) - C_q (*sdhA*, assay strain)
$\Delta C_q 3$	C_q (*gfpmut3*, BAC0297) - C_q (*gapA*, BAC0297)
$\Delta C_q 4$	C_q (*gfpmut3*, BAC0297) - C_q (*sdhA*, BAC0297)

The normalised target gene transcription level was calculated for the assay and control strains using Equation 2-4, which assumes 100% product amplification efficiency. Lastly the normalised *gfpmut3* transcription level was calculated relative to *gapA* using Equation 2-5, and *sdhA* using Equation 2-6 ($\Delta\Delta C_q$). The standard deviation of the average final relative transcription value, between the biological replicates, was calculated.

Equation 2-4 – Normalised gfpmut3 transcription level

$$\mathbf{2}^{-\Delta Cq}$$

Equation 2-5 – Normalised gfpmut3 transcription level relative to gapA

$$\frac{\mathbf{2}^{-\Delta Cq1}}{\mathbf{2}^{-\Delta Cq3}}$$

Equation 2-6 – Normalised gfpmut3 transcription level relative to sdhA

$$\frac{\mathbf{2}^{-\Delta Cq2}}{\mathbf{2}^{-\Delta Cq4}}$$

2.15 CRISPR enabled DNA base editing in *Bacillus subtilis*

Plasmids harbouring dCas9-AID, nCas9-AID, TadA-dCas9 and TadA-nCas9 were used to transform naturally competent *B. subtilis* (Section 2.4.2) and successful transformants were identified on LB agar plates supplemented with chloramphenicol and 1 mmol L^{-1} IPTG. dCas9 and nCas9 plasmids targeting the same region were used as controls to show base editing was caused by the fused deaminase.

2.15.1 TadA base editing

cPCR and sequencing was carried out of the target region within transformants. Three transformants were used to inoculate 5 mL LB broth supplemented with chloramphenicol and 1 mmol L^{-1} IPTG and grown to exponential growth phase at 37 °C, 250 rpm. This was used to inoculate 10 mL LB broth supplemented with chloramphenicol and 1 mmol L^{-1} IPTG to OD_{600} 0.05 and grown for 48 hours at 37 °C, 250 rpm. At 16, 24 and 48 hours of growth, 10 µL of this culture was struck only on an LB agar plate supplemented with chloramphenicol only and grown overnight. Colonies were analysed by cPCR and sequenced to determine if base editing had successfully occurred at the target region.

2.15.2 AID base editing

2.15.2.1 AID base editing of a single target nucleotide

cPCR and sequencing of the transformants identified a mixed population of edited and non-edited cells in analysed colonies. Transformants were struck out to obtain single colonies on LB agar supplemented with chloramphenicol only. The proportion of the original transformation colony with successful base editing was established by colony PCR and sequencing of this subpopulation.

2.15.2.2 AID base editing of multiple target nucleotides

Triplicate transformants were used to inoculate 5 mL LB broth supplemented with chloramphenicol and 1 mmol L^{-1} IPTG and grown to exponential growth phase at 37 °C, 250 rpm. This was used to inoculate 10 mL LB broth supplemented with chloramphenicol and 1 mmol L^{-1} IPTG to OD_{600} 0.05 and grown for 24 hours at 37 °C, 250 rpm. 10 µL of this culture was struck only on an LB agar plate supplemented with chloramphenicol and 1 mmol L^{-1} IPTG and grown overnight. Colonies were analysed by cPCR and sequenced to determine if base editing had successfully occurred.

2.16 *In silico* data analysis

All data calculations were carried out using Microsoft Excel (Microsoft, USA). This data was subsequently represented in graphs using GraphPad Prism 6 (GraphPad Software, USA).

In silico DNA manipulations were performed using Clone Manager 9 (SciEd Software, USA). Following DNA sequencing, alignment with reference sequences was performed using Clone Manager 9 (SciEd Software, USA) while DNA chromatograms were interpreted using the ApE – A plasmid Editor. Vector schemes were prepared in Vector NTI® (Thermo Fisher Scientific, USA).

2.17 *In silico* CRISPR array library construction

To determine the theoretical maximum number of inserts for the construction of a CRISPR array library, the Edinburgh Genome Foundry CUBA software was utilised. Specifically, the design overhangs tool allowed the identification of compatible overhangs within the *S. pyogenes* CRISPR array direct repeat region.

Chapter 3 – CRISPR-enabled genome editing in *Bacillus subtilis*

3.1 Introduction

CRISPR-Cas9 systems of adaptive immunity in bacteria have become widely used across all fields of biology as a genome editing tool since demonstration of its use as a RNA-guided DNA endonuclease in 2012[43]. This method employs the Cas9 endonuclease being guided to a target region, neighbouring a 5´-NGG-3´ protospacer adjacent motif (PAM) site, by an sgRNA consisting of a targeting protospacer region, the 'handle' to which the Cas9 binds (tracrRNA when not fused to crRNA within an sgRNA), and a stem loop combining these features[61]. The technique, based on the type II CRISPR-Cas9 system from *Streptococcus pyogenes*, makes use of the host DNA double-strand break (DSB) repair machinery to introduce mutations within the DNA sequence. The repair of these DSBs, generally by homology directed repair (HDR) mechanisms, allows the introduction of selected modifications in the presence of an engineered template with chromosome homology regions flanking the desired mutations.

Following the establishment of the CRISPR-Cas9 system in *B. subtilis* in 2016, it was striking that the initial methods employed involved the editing of a strain harbouring a chromosomal integration of *cas9*, followed by a co-transformation of an sgRNA expression cassette and dDNA (both of which would be integrated onto the chromosome)[105,135]. Chromosomal integration of *cas9* and sgRNA expression cassettes were utilised since the use of plasmids within *B. subtilis* is lower than other hosts, such as *E. coli*, due to their historically comparative instability. Such instabilities have been overcome however, in part through the incorporation of *repA* within plasmids such as pHT01 (MoBiTec) allowing efficient theta plasmid replication[136]. These systems are burdened with an inflexibility to their wide scale implementation due to the additional work required to validate the insertion of *cas9*. A further example of the CRISPR-Cas9 system, developed by Altenbuchner, employed the use of a single plasmid for chromosomal editing, showing that prior chromosomal integration of *cas9* was not essential[63]. This plasmid harboured the *cas9* gene, sgRNA expression cassette, and dDNA. While the flexibility of this system is increased relative to the chromosomal integration system, the need for

both sgRNA and dDNA plasmid cloning increases the time for experiments to be completed.

To address these concerns, this chapter describes the development of a system, based on the *B. subtilis* – *E. coli* shuttle vector, pHT01, where sgRNA and *cas9* are expressed on a single plasmid. This final plasmid, used alongside a separate PCR product as dDNA for *B. subtilis* co-transformation, allows chromosomal modifications through either homologous recombination (HR) or HDR. Cas9 is unable to cleave the target site following successful chromosomal integration of the dDNA due to a synonymous PAM site mutation preventing further cleavage. The editing plasmid is then subsequently removed from the cell, leaving the successfully edited strain for downstream applications.

The use of PCR products as dDNA was shown by Westbrook *et al.* to be less efficient for transformations compared to when the dDNA was contained on a separate plasmid[135]. However, the flexibility of PCR products as dDNA (compared to the time and resources required for plasmid construction), combined with the increased flexibility of a plasmid based *cas9* and sgRNA expression system, rather than chromosomal integration, was preferable and is used in this study. Furthermore, the reliably high transformation efficiencies obtained with the adapted naturally competent *B. subtilis* 168 method, developed by Bennallack *et al*., ensures that a co-transformational approach is viable[24].

Due to the known high rate of homologous recombination in *B. subtilis*, it was hypothesised that the primary mode of action for CRISPR-Cas9 enabled genomic modifications was due to homologous recombination occurring first, with the CRISPR-Cas9 system acting as a counterselection method for these recombination events. It has remained unclear if the Cas9 induced DSB induces DNA repair. To test this hypothesis, the tryptophan auxotrophy of *B. subtilis* 168 was exploited. This auxotrophy is caused by the absence of a single isoleucine codon in the *trpC2* gene[63]. The reintroduction of this codon restores prototrophy and enables growth on M9 minimal medium without the addition of L-tryptophan.

A consideration for commercial applications of CRISPR-Cas9 has been the intellectual property (IP) landscape surrounding its use. While the IP landscape for use in bacteria appears to be relatively clear when compared to its use in eukaryotes and medical applications, the requirement for licence and royalty payments remains a considerable disincentive for commercial use. An alternative

CRISPR nuclease family, known as Cpf1 or Cas12a, were initially identified in 46 different organisms, with 16 being characterised for PAM specificities and 8 found to efficiently induce DNA DSB[137]. These nucleases utilise alternative mechanisms to identify their targets and induce DNA cleavage. T rich PAM sites were characterised as well as the DNA being cut with a 5-8 base staggered cut rather than a blunt cut as performed by Cas9[137]. Patents covering the use of Cpf1 were quickly filed following its discovery, thus making its use similarly unattractive.

The release of a further alternative CRISPR nuclease, MAD7 (isolated from *Eubacterium rectale* and also known as ErCas12a[71]), in 2018 represented a potential way to bypass the use of the Cas9 or Cpf1 nucleases. The developers of this nuclease, Inscripta (USA), claim MAD7 is free to use for both academic and industrial research applications. They also state that any strain developed using MAD7 is royalty free for commercial applications, provided that the final strain does not contain the MAD7 nuclease itself[70]. While Inscripta have released data confirming MAD7 as active for genomic editing in *E. coli*, *S. cerevisiae,* mouse and human HEK293T cells, and Wesley *et al*. recently exemplified its activity in zebrafish, data on its activity levels in *B. subtilis* was lacking[71].

In this chapter the CRISPR-Cas9 method outlined above is developed, as well as comparing this system with the alternative CRISPR nucleases, *Acidaminococcus sp.* BV3L6 Cpf1 (AsCpf1) and MAD7. Furthermore, the driving mechanism behind CRISPR genome editing within *B. subtilis* 168 was investigated and a licence free protease knock-out library is designed and partially constructed.

The work in this chapter relating to MAD7 was published in February 2020 and a copy of this publication can be found in Appendix G.

3.2 Materials and methods

3.2.1 Plasmids and strains

The plasmids and strains used in this chapter are outlined below in Table 3.1. The oligonucleotides used in this chapter can be found in Appendix C.

Table 3.1 – Strains and plasmids used in this chapter.

Strain/Plasmid	Description/Genotype	Reference
Strains		
B. subtilis 168	*trpC2*	Laboratory stock
BAC0040	*B. subtilis* 168 with pBAC0008.	This chapter.
BAC0050	*B. subtilis* 168 with pBAC0015.	This chapter.
BAC0111	*B. subtilis* 168 *ΔbglS::aph(3´)-lIIa*-LIC site-*gfpmut3*.	This chapter.
BAC0288	BAC0111 *ΔbglS::aph(3´)-lIIa*-P_{veg}-*gfpmut3*.	This chapter.
BAC0369	*B. subtilis* 168 *ΔaprE*	This chapter.
BAC0371	*B. subtilis* 168 *ΔnprE*	This chapter.
BAC0371	*B. subtilis* 168 *ΔhtrA*	This chapter.
BAC0372	*B. subtilis* 168 *Δbpr*	This chapter.
BAC0373	*B. subtilis* 168 *Δmpr*	This chapter.
BAC0374	*B. subtilis* 168 *Δepr*	This chapter.
Plasmids		
pHT01	*E. coli/B. subtilis* shuttle vector carrying P_{grac} and *lacI, bla*, *cat*.	MoBiTec
pGFPbglS	*bla*; 5´ *bglS*; *aph(3´)-lIIa*; LIC site; *gfpmut3*; 3´ *bglS*	[124]
pMK-RQ-MAD7	ColE1 ori; *aphAI*; *MAD7* codon optimised for *B. subtilis.*	Synthesised by Thermo Fisher Scientific

pMK-RQ-gfpmut3-dDNA	ColE1 ori; *aphAI*; *gfpmut3* dDNA for stop codon introduction.	Synthesised by Thermo Fisher Scientific
pY010 (pcDNA3.1-hAsCpf1)	ColE1 ori*; bla*; P_{CMV}; P_{T7}; *AsCpf1*-NLS; f1 ori; *aph(3´)-II.*	[137]
pBAC0001	pHT01 with SapI sites removed.	This chapter.
pBAC0003	pBAC0001 with rrnB T1 T2; sgRNA Cas9 handle; protospacer cloning site; $P_{J23110.}$	This chapter.
pBAC0008	pBAC0001 with rrnB T1 T2; sgRNA Cas9 handle; protospacer cloning site; $P_{veg.}$	This chapter.
pBAC0009	pBAC0001 with rrnB T1 T2; sgRNA Cas9 handle; protospacer cloning site; $P_{rrnB\ P1.}$	This chapter.
pBAC0015	pBAC0008 with *cas9* (*S. pyogenes*). Cas9 expression regulated by the P_{grac}.	This chapter.
pBAC0027	pBAC0015 with sgRNA targeting 147 bp downstream of the start codon within *B. subtilis* 168 *amyE*.	This chapter.
pBAC0035	pBAC0015 with a non-targeting protospacer.	This chapter.
pBAC0041	pBAC0015 with sgRNA targeting 25 bp downstream of the start codon within *B. subtilis* 168 *amyE*.	This chapter.
pBAC0047	pBAC0015 with sgRNA targeting 83 bp downstream of the start codon within *B. subtilis* 168 *amyE*.	This chapter.
pBAC0085	pBAC0015 with sgRNA targeting 347 bp downstream of the start codon within *B. subtilis* 168 *trpC2*.	This chapter.
pBAC0129	pBAC0015 with sgRNA targeting 18 bp upstream of the start codon within BAC0111 *gfpmut3*.	This chapter.

pBAC0154	pBAC0001 with rrnB T1 T2; crRNA AsCpf1 handle; protospacer cloning site; P_{veg}	This chapter.
pBAC0155	pBAC0001 with rrnB T1 T2; crRNA MAD7 handle; protospacer cloning site; P_{veg}	This chapter.
pBAC0158	pBAC0155 with *MAD7* cloned between BamHI and XbaI sites. The *MAD7* gene was isolated from BamHI and XbaI restriction digested pBAC0145.	This chapter.
pBAC0159	pBAC0154 with *Ascpf1* cloned between BamHI and XbaI sites. The *Ascpf1* gene was isolated from pBAC0146 using oligonucleotides oMAP0500/0501.	This chapter.
pBAC0160	pBAC0159 with a non-targeting gRNA.	This chapter.
pBAC0161	pBAC0159 with gRNA targeting 21 bp downstream of the start codon within *B. subtilis* 168 *amyE*.	This chapter.
pBAC0162	pBAC0158 with gRNA targeting 21 bp downstream of the start codon within *B. subtilis* 168 *amyE*.	This chapter.
pBAC0163	pBAC0158 with a non-targeting gRNA.	This chapter.
pBAC0165	pBAC0015 with sgRNA targeting 27 bp downstream of the start codon within BAC0288 *gfpmut3*.	This chapter.
pBAC0166	pBAC0158 with gRNA targeting 21 bp downstream of the start codon within BAC0288 *gfpmut3*.	This chapter.
pBAC0167	pBAC0159 with gRNA targeting 21 bp downstream of the start codon within BAC0288 *gfpmut3*.	This chapter.
pBAC0185	pBAC0015 with sgRNA targeting 368 bp downstream of the start codon within *B. subtilis* 168 *trpC2*.	This chapter.
pBAC0197	pBAC0158 with gRNA targeting 12 bp downstream of the start codon within *B. subtilis* 168 *aprE*.	This chapter.
pBAC0198	pBAC0158 with gRNA targeting 6 bp downstream of the start codon within *B. subtilis* 168 *nprE*.	This chapter.

pBAC0199	pBAC0158 with gRNA targeting 22 bp downstream of the start codon within *B. subtilis* 168 *htrA*.	This chapter.
pBAC0200	pBAC0158 with gRNA targeting 25 bp downstream of the start codon within *B. subtilis* 168 *htrB*.	This chapter.
pBAC0201	pBAC0158 with gRNA targeting 1 bp upstream of the start codon within *B. subtilis* 168 *bpr*.	This chapter.
pBAC0202	pBAC0158 with gRNA targeting 2 bp upstream of the start codon within *B. subtilis* 168 *nprB*.	This chapter.
pBAC0203	pBAC0158 with gRNA targeting 1 bp upstream of the start codon within *B. subtilis* 168 *mpr*.	This chapter.
pBAC0204	pBAC0158 with gRNA targeting 6 bp downstream of the ATG start codon within *B. subtilis* 168 *epr*.	This chapter.
pBAC0205	pBAC0158 with gRNA targeting 15 bp downstream of the start codon within *B. subtilis* 168 *wprA*.	This chapter.
pBAC0206	pBAC0158 with gRNA targeting 7 bp downstream of the start codon within *B. subtilis* 168 *vpr*.	This chapter.
pBAC0218	pBAC0158 with gRNA targeting 299 bp downstream of the start codon within *B. subtilis* 168 *trpC2*.	This chapter.

3.2.1.1 Plasmid construction

pBAC0001 was constructed by the Gibson DNA assembly method (Section 2.5.11) to remove the SapI sites present in pHT01 (MoBiTec) to allow compatibility with the inABLE® DNA assembly method[122]. PCR products (Section 2.5.5) for Gibson assembly were prepared from pHT01 using oligonucleotides oMAP0002/0003/0004/0005.

Plasmids pBAC0003/0008/0009/0154/0155 were constructed using the inABLE® DNA assembly technique (Section 2.5.12).

pBAC0008, designed for use in experiments where *S. pyogenes* Cas9, or its derivatives, would be used, consisted of four parts: 1. the *E. coli/B. subtilis* shuttle

vector backbone from pBAC0001; 2. the LacI repressor and isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible P_{grac} promoter from pBAC0001, including a multiple cloning site; 3. the bidirectional strong rrnB T1 and T2 terminators; 4. the sgRNA expression module consisting of a kanamycin resistance gene (*aphAI*) flanked by AarI sites expressed under the control of the P_{veg} promoter, and the 'Cas9 handle' section of the sgRNA. 5´ truncated parts were prepared by PCR from the indicated template and oligonucleotides: 1. pBAC0001 with oMAP0010/0011; 2. pBAC0001 with oMAP0018/0019; 3. pING0001 with oMAP0024/0025; 4. pING0002 with oMAP0030/0031. Parts were ligated at 5´ and 3´ respectively with annealed oligonucleotides: 1. oMAP0008/0009 and oMAP0014/0015; 2. oMAP0016/0017 and oMAP0020/0021; 3. oMAP0022/0023 and oMAP0048/0049; 4. oMAP0050/0051 and oMAP0052/0053.

pBAC0003 and pBAC0009 plasmids were identical to pBAC0008 with the exception that P_{veg} was replaced by P_{J23110} and $P_{rrnB\ P1}$ respectively for sgRNA expression. 5´ truncated parts were prepared in the same manner. pBAC0003/0009 parts 1., 2. and 3. were ligated with the same sets of annealed oligonucleotides as pBAC0008. The remaining part 4. for pBAC0003 was ligated at 5´ and 3´ respectively with annealed oligonucleotides oMAP0050/0051 and oMAP0012/0013 to introduce P_{J23110}. The remaining part 4. for pBAC0009 was ligated at 5´ and 3´ respectively with annealed oligonucleotides oMAP0050/0051 and oMAP0054/0055 to introduce $P_{rrnB\ P1}$.

pBAC0154, designed for use in experiments where *Acidaminococcus* Cpf1, or its derivatives, would be used, consisted of four parts: 1., 2. and 3. were the same corresponding parts as used in pBAC0008; 4. gRNA expression module consisting of non-coding spacer DNA flanked by SapI sites and the 'AsCpf1 handle' section of the gRNA under the control of the P_{veg} promoter. Due to its size, this part was prepared using only phosphorylated and annealed oligonucleotides (oMAP0490/0491), thus allowing SapI sites to be introduced into the part for later cloning steps (Parts were ligated at 5´ and 3´ respectively with annealed oligonucleotides: 1. and 2. as in pBAC0008; 3. oMAP0022/0023 and oMAP0486/0487; 4. oMAP0488/0489 and oMAP0498/0499.).

pBAC0155, designed for use in experiments where MAD7 (Inscripta Inc., USA), or its derivatives, would be used, consisted of four parts: 1., 2. and 3. were the same corresponding parts as used in pBAC0008; 4. gRNA expression module

consisting of non-coding spacer DNA flanked by SapI sites and the 'Mad7 handle' section of the gRNA under the control of the P_{veg} promoter. Due to its size, this part was prepared using only phosphorylated and annealed oligonucleotides (oMAP0492/0493), thus allowing SapI sites to be introduced into the part for later cloning steps (Parts were ligated at 5´ and 3´ respectively with annealed oligonucleotides: 1. and 2. as in pBAC0008; 3. oMAP0022/0023 and oMAP0486/0487; 4. oMAP0488/0489 and oMAP0498/0499).

pBAC0013 was constructed by converting the catalytically inactive *dcas9* gene from pdCas9-bacteria to active *cas9*. This was done by introducing mutations A10D, A840H and removing a BamHI site by PCR (oligonucleotides oMAP0062/0063/0064/0065/0066/0067) and a subsequent 3-part Gibson Assembly (pdCas9-bacteria was a gift from Stanley Qi (Addgene plasmid # 44249))[122,138].

The catalytically active *cas9* gene was PCR amplified from pBAC0013 with primers oMAP0073/0074, introducing a BsaI site and XbaI restriction enzyme recognition sites at the 5´ and 3´ end of the gene, respectively. The amplified *cas9* was digested (Section 2.5.8) with BsaI-HF and XbaI, and ligated (Section 2.5.10) with the BamHI-HF and XbaI digested pBAC0008 backbone, yielding pBAC0015. BsaI was used due to WT *S. pyogenes cas9* containing a BamHI recognition site.

To analyse the alternative CRISPR nucleases, MAD7, pBAC0158 was constructed by digesting pBAC0155 and pMK-RQ-MAD7 (*MAD7*, codon optimised for *B. subtilis* and flanked by BamHI and XbaI recognition sites, synthesised by Thermo Fisher Scientific) with BamHI-HF and XbaI restriction enzymes, and ligating the pBAC0155 backbone with the *MAD7* gene.

pBAC0159 was constructed by PCR amplifying *AsCpf1* from pY010 (pcDNA3.1-hAsCpf1) with oligonucleotides oMAP0500/0501, digesting this and pBAC0154 with BamHI-HF and XbaI restriction enzymes and ligating the pBAC0154 backbone with the *AsCpf1* gene. pY010 (pcDNA3.1-hAsCpf1) was a gift from Feng Zhang (Addgene plasmid # 69982)[137].

Plasmids used for Cas9, AsCpf1 or MAD7-mediated editing were prepared from the pBAC0015, pBAC0159 or pBAC0158 parental plasmids respectively, using phosphorylated and annealed oligonucleotide pairs as described in Section 2.5.4.1. Information on the oligonucleotides used in construction, target gene and PAM site for the following plasmids used in this chapter can be found in Table 2.12: pBAC0027; pBAC0041; pBAC0047; pBAC0085; pBAC0129; pBAC0161;

pBAC0162; pBAC0165; pBAC0166; pBAC0167; pBAC0185; pBAC0197; pBAC0198; pBAC0199; pBAC0200; pBAC0201; pBAC0202; pBAC0203; pBAC0204; pBAC0205; pBAC0206; pBAC0218.

As a positive control for transformation efficiency, an sgRNA designed not to target the *B. subtilis* 168 chromosome was inserted into the AarI digested and dephosphorylated pBAC0015 backbone with the phosphorylated and annealed oligonucleotide pair oMAP0145/0147, yielding pBAC0035. Similar non-targeting gRNAs were inserted into the SapI digested and dephosphorylated pBAC0159 (oMAP0553/0554) and pBAC0158 (oMAP0555/0556) yielding pBAC0160 (AsCpf1) and pBAC0163 (MAD7) transformation positive control plasmids respectively.

3.2.1.2 Strain construction

BAC0111 was constructed by transforming naturally competent *B. subtilis* 168 (Section 2.4.2) with the integration plasmid pGFPbglS[16,24,124]. Transformants were selected on LB agar plates supplemented with kanamycin.

Using the approach described in section 2.7, CRISPR-Cas9 mediated chromosome editing was used to insert the P_{veg} promoter upstream of the *gfpmut3* gene in BAC0111 to construct strain BAC0288. Linear dDNA in the form of overlap extension PCR (OE-PCR) products, amplified from BAC0111 gDNA with oligonucleotide sets oMAP0388/0393/0394/0395 (Section 2.5.5.3) was used alongside pBAC0129. The CRISPR-Cas9/sgRNA plasmid was removed from the edited strain as described in Section 2.7.6. The insertion of P_{veg} was verified by PCR amplification from BAC0288 gDNA with oMAP0803/0804 and sequencing. Furthermore, GFPmut3 was confirmed by fluorescence emission analysis using Safe Imager 2.0 Blue Light Transilluminator and Amber Filter (Thermo Fisher Scientific) system (excitation: 470 nm, emission: 530 nm).

A single OE-PCR product was used to allow a direct comparison between Cas9, AsCpf1 and MAD7 editing efficiencies at the *amyE* locus. The OE-PCR product generated using the oligonucleotide set oMAP0121/0551/0552/0122 was co-transformed alongside the editing plasmids pBAC0041, pBAC0161 and pBAC0162 for Cas9, AsCpf1 and MAD7 editing respectively. Similarly, a single dDNA was generated for *gfpmut3* editing in conjunction with each nuclease. This was generated by PCR, using oligos oMAP0575/0578 and the synthesised plasmid

pMK-RQ-gfpmut3-dDNA as template, and transformed alongside the editing plasmids pBAC0165, pBAC0167 and pBAC0166 for Cas9, AsCpf1 and MAD7 editing respectively. When targeting *gfpmut3*, transformants were spread on LB agar plates supplemented with chloramphenicol and IPTG (1 mmol L^{-1}). Effective knock-out of *gfpmut3* by stop codon introduction was determined by analysis of fluorescence emission using Safe Imager 2.0 Blue Light Transilluminator and Amber Filter (Thermo Fisher Scientific) system (excitation: 470 nm, emission: 530 nm). Genotypes were confirmed by colony PCR with oligonucleotides hybridising to the chromosome (oMAP0393/0814) outside of the dDNA homology arm region, and sequenced with primer oMAP0815 to ensure accurate coverage of the targeted region. When targeting *amyE*, transformants were spread on LB agar plates supplemented with chloramphenicol, IPTG (1 mmol L^{-1}) and 1% w/v soluble potato starch (VWR). Effective knock-out of *amyE* by stop codon introduction was determined by staining transformation plates with iodine[1,27]. Genotypes were confirmed by colony PCR with oligonucleotides hybridising to the chromosome (oMAP0811/0812) outside of the dDNA homology arm region, and sequenced with primer oMAP0813 to ensure accurate coverage of the targeted region.

To construct the protease knock-out strain library, 200 ng of plasmids pBAC0197-0206 (targeting protease encoding genes outlined in Table 3.1) were co-transformed with 1 µg of the appropriate dDNA in the form of a OE-PCR product prepared from *B. subtilis* 168 gDNA with the following oligonucleotide sets: oMAP0198/0703/0704/0195 (pBAC0197), oMAP0255/0707/0708/0252 (pBAC0198), oMAP0261/0711/0712/0258 (pBAC0199), oMAP0267/0715/0716/0264 (pBAC0200), oMAP0726/0727/0728/0729 (pBAC0201), oMAP0733/0734/0735/0736 (pBAC0202), oMAP0740/0741/0742/0743 (pBAC0203), oMAP0747/0748/0749/0750 (pBAC0204), oMAP0754/0755/0756/0757 (pBAC0205), oMAP0719/0720/0721/0722 (pBAC0206).

3.3 Results

3.3.1 Construction and assessment of a single-plasmid approach for CRISPR-Cas9-mediated gene editing in *B. subtilis*

To ensure the plasmid construction for final editing plasmids is as simple as possible, a single plasmid editing approach was developed which allows rapid *in situ* edits of the genome, using a plasmid which can target any *loci* within *B. subtilis* in a

single cloning step, and without previous genome modifications. During the editing process, this plasmid transcribes a sgRNA and *cas9*. The dDNA is co-transformed with the plasmid in the form of a linear PCR product constructed using overlap extension PCR (OE-PCR). The dDNA includes the desired genome mutations as well as a synonymous mutation to remove the PAM site and prevent further cutting by the nuclease. In each dDNA, mutations are flanked either side by regions of ~1 kbp homologous to the chromosome either side of the targeted PAM site. CRISPR-Cas9 editing plasmids were designed based on the commercially available *E. coli* - *B. subtilis* shuttle vector, pHT01.

3.3.1.1 Promoter selection for sgRNA expression

Accuracy of the sgRNA transcriptional start site is important for high editing efficiency. This is due to the 5´ region of the sgRNA being the 20 nt protospacer region and errors within the transcriptional start site can reduce the capacity for targeting Cas9 to the relevant site[61]. It was decided that constitutive promoters would be used for sgRNA expression to ensure an excess of sgRNA and prevent this being a limiting factor, maximising editing efficiency. Three constitutive promoters (synthetic P_{J23101}(constitutive), and native P_{veg} (constitutive) and $P_{rrnB\ P1}$ (active during exponential growth phase)) with well characterised transcriptional start sites had previously been identified[139,140] and had the additional benefit of being short enough to be included on oligonucleotides used during plasmid preparation with the inABLE® approach (section 3.2.1.1).

To ensure these promoters had sufficient activity for sgRNA expression, the kanamycin resistance *aphAI* gene was introduced between AarI restriction enzyme recognition sites within pBAC0003 (P_{J23101}), pBAC0008 (P_{veg}) and pBAC0009 ($P_{rrnB\ P1}$). *B. subtilis* 168 was transformed with each plasmid, with successful transformants being selected on LB agar supplemented with chloramphenicol. Subsequently, three transformants were struck onto LB agar supplemented with chloramphenicol and kanamycin. Growth was observed with all colonies containing the pBAC0008 and pBAC0009 plasmids, but not pBAC0003. As such, P_{J23101} was judged to be insufficiently active for sgRNA expression. $P_{veg,}$ was selected to be taken forward for sgRNA expression due to it being more commonly used for expression within *B. subtilis*[118]. While P_{veg} is natively 237 bp in length, Sojka *et al*. (2011) described a shortened version of 42 bp lacking one of the two σ^A binding

sites to give a medium rather than strong promoter strength which was used in this study[140].

3.3.1.2 Validation that cas9 is expressed to allow editing

Since *S. pyogenes* Cas9 is the most widely reported and well characterised CRISPR nuclease to date it was utilised throughout this study[141]. Due to the reported toxicity of Cas9 to some organisms, it was decided to utilise the IPTG inducible P_{grac} promoter (present in pHT01) for *cas9* expression which would allow the tuning of expression levels should it be required[142,143].

Cas9 was cloned into pBAC0008 as described in section 3.2.1.1, yielding pBAC0015. Following transformation of *B. subtilis* 168 with plasmids pBAC0008 and pBAC0015, producing strains BAC0040 and BAC0050 respectively, protein expression analysis was performed. The strains were grown to a normalised cell state to ensure comparability between cultures (Section 2.2.1) and subsequently grown with and without IPTG for 18 hours before harvesting, preparing samples and performing SDS-PAGE analysis (Section 2.8). The gel was stained with Instant Blue™ (Section 2.8.3) and scanned (Figure 3.1).

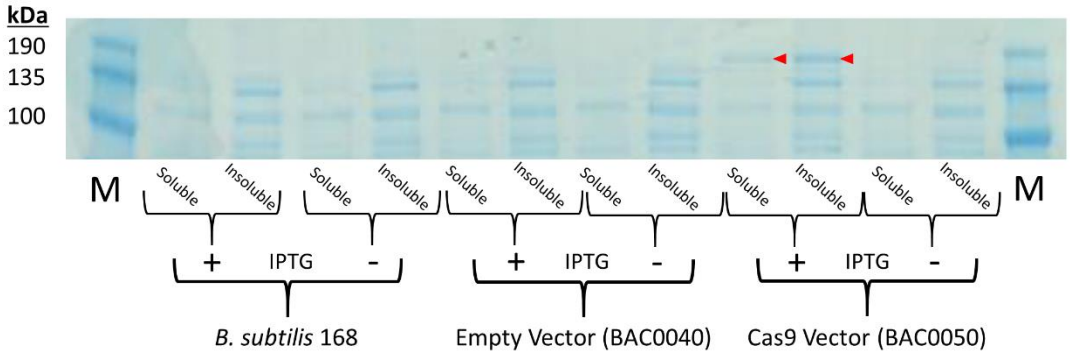


Figure 3.1 - *cas9* expression SDS-PAGE analysis.

SDS-PAGE gel analysis of strains *B. subtilis* 168, BAC0040 (empty vector control) and BAC0050 (pBAC0015 for *cas9* expression) for the confirmation of *cas9* expression. Soluble and insoluble expression bands are shown following chemical cell lysis (Section 2.8.1). Strains were grown in the presence (+) or absence (-) of IPTG. Protein size was determined by comparison Blue pre-stained protein standard, broad range (M). The predicted band size for Cas9 was 158 kDa, indicated by the red arrows.

Cas9 (158 kDa) expression can be observed in BAC0050 in the presence of IPTG, with the majority of expression within the insoluble fraction. No other obvious differences were observed between the empty vector control strain, and the Cas9 containing strain.

3.3.1.3 Cloning strategy for the simple, rapid completion of CRISPR-Cas9 editing plasmids

To complete the cloning of each editing plasmid, the parental plasmid, such as pBAC0015, requires the cleavage of two type IIS restriction enzyme sites yielding non-compatible overhangs (Figure 3.2). This removes the *aphAI* gene which confers the kanamycin resistance used to identify suitable sgRNA promoters (section 3.3.1.1). Furthermore, the removal of the 858 bp DNA fragment between the AarI sites acts as an indicator to successful DNA cleavage. To ensure no self-ligation in cases where only one of the two AarI sites has successfully been cleaved, the digested DNA is dephosphorylated (Section 2.5.9).

A 20 bp DNA fragment is inserted to complete the sgRNA. The DNA for insertion is simply prepared by the phosphorylation and annealing of two 24 nt DNA oligonucleotides with compatible overhangs to the cut plasmid (Figure 3.2C). The subsequent completed CRISPR plasmid, once confirmed by sequencing with oMAP0148, is ready for genome editing to proceed. This system represents a simple, rapid approach to target Cas9 to any site within the genome provided a suitable PAM site is present.

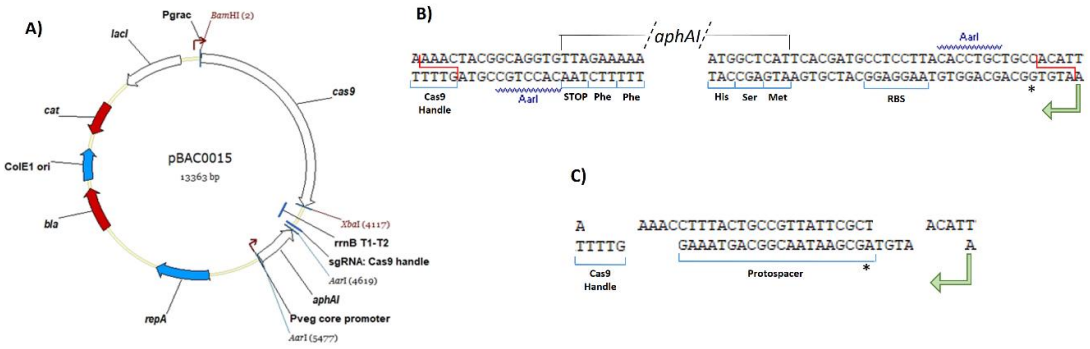


Figure 3.2 – Cloning strategy for final CRISPR-Cas9 editing plasmids.

A) Schematic of the 'parental' CRISPR-Cas9 editing plasmid, pBAC0015. **B)** Protospacer cloning region prior to digestion with the AarI restriction enzyme. AarI recognition sites are indicated and the corresponding cut sites and exposed ssDNA overhangs are indicated by red lines. The ribosome binding site (RBS) and amino acid sequence for the N-terminus and C-terminal ends of the kanamycin resistance gene, *aphAI*, are indicated within the blue brackets. The 5´ region of the downstream Cas9 Handle are also represented. The 3' region of P_{veg} is indicated by the green arrow. * = transcriptional start site for P_{veg}. **C)** Protospacer cloning region post AarI digestion. 4 nt ssDNA overhangs are exposed, allowing cloning to proceed. Two phosphorylated and annealed complimentary oligonucleotides (harbouring the protospacer to guide Cas9 to its target site) with 4 nt 5´ ssDNA overhangs with complementarity to the AarI digested and dephosphorylated pBAC0015 backbone and subsequently be ligated to complete the cloning procedure.

3.3.1.4 Assessment of the CRISPR-Cas9-enabled genome editing efficiency within *B. subtilis* 168

In order to determine the efficiency of genome editing with the developed system, three PAM sites were identified at the 5´ end of the gene encoding the starch degrading α-Amylase enzyme, *amyE*. This was chosen due to the simple identification of knock-out mutants following growth on media containing starch. When starch is exposed to iodine, a deep blue colour is obtained. Since this enzyme is secreted into the extracellular environment, when agar containing soluble starch is stained with iodine, the absence of a halo where a colony was formed indicates the absence of secreted α-Amylase. Thus, through the introduction of stop codons at the 5´ end of the gene to disrupt translation of the mRNA, knock-out mutants could be identified which correlate directly to editing efficiency.

These three protospacers were cloned into pBAC0015 (as per the approach outlined in Figure 3.2) yielding plasmids pBAC0027/0041/0047 (Table 2.12). The cloning of the protospacers was confirmed by colony PCR, using the sense strand protospacer oligonucleotide as the forward primer and a reverse primer hybridising in the plasmid backbone (oMAP0148) and sequencing of the plasmid DNA. An example of such a colony PCR reaction is shown in Figure 3.3 where 2 of 3 clones screened successfully had the protospacer inserted.

dDNA for use with pBAC0027/0041/0047 was developed by OE-PCR using the oligonucleotide sets oMAP0121/0122/0123/0124, oMAP0121/0122/0128/0129 and oMAP0121/0122/0143/0144 respectively. Each dDNA was designed to introduce a stop codon together with a synonymous PAM mutation to eliminate Cas9 cleavage at the edited site after homologous recombination. Naturally competent *B. subtilis* 168 was co-transformed with the plasmid/dDNA pairs and grown in the presence or absence of IPTG to determine if leaky expression of *cas9* from P_{grac} was sufficient for genome editing and if editing efficiency could be increased with a higher level of nuclease present. The non-targeting plasmid, pBAC0035 was transformed alongside these experiments as a control to show cleavage would only take place in the presence of the sgRNA and nuclease.

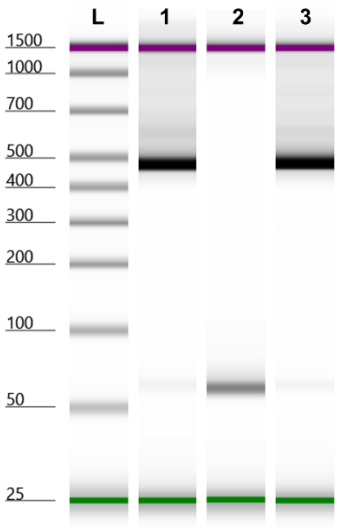


Figure 3.3 – Capillary gel electrophoresis following colony PCR to detect clones with successful insertion of protospacer DNA.

D1000 ScreenTape analysis with the Agilent 2200 TapeStation system (Section 2.5.6.2). DNA band sizes (bp) were compared to an electronic ladder (L). 1 – clone 1; 2 – clone 2; 3 – clone 3. The expected band size for a successful protospacer insertion was 444 bp.

The results (Figure 3.4) show effective knock-out colonies are only obtained in the presence of dDNA. In the absence of IPTG, an editing rate for pBAC0027/0041/0047 of 89.3%, 89.5% and 82.4% was achieved respectively. In the presence of IPTG the rates were 83.8%, 85.3% and 86.8% indicating that higher levels of nuclease expression do not have a significant effect on editing rate. This suggests that the rate of editing is independent of Cas9 levels, with only a requirement of basal 'leaky' expression for counterselection by cleaving and preventing growth of the non-edited cells. While knock-out colonies could have been obtained through the use of dDNA only, the lack of a selection mechanism would cause the formation of a lawn following transformation in the absence of antibiotic selection for plasmid incorporation. The level of editing in the absence of antibiotic selection and Cas nuclease counterselection is discussed more in section 3.3.3. Additionally, the base level editing efficiency with antibiotic selection for plasmid incorporation but without Cas nuclease counterselection is investigated in section 3.3.2.

Since no deletion phenotypes were identified in the absence of dDNA, we have shown that CRISPR-Cas9 editing does not occur due to NHEJ in *B. subtilis* 168. Instead we have shown that editing is bound to either HDR or HR (with Cas9 acting as an efficient counterselection tool), which is further investigated below in section 3.3.3.

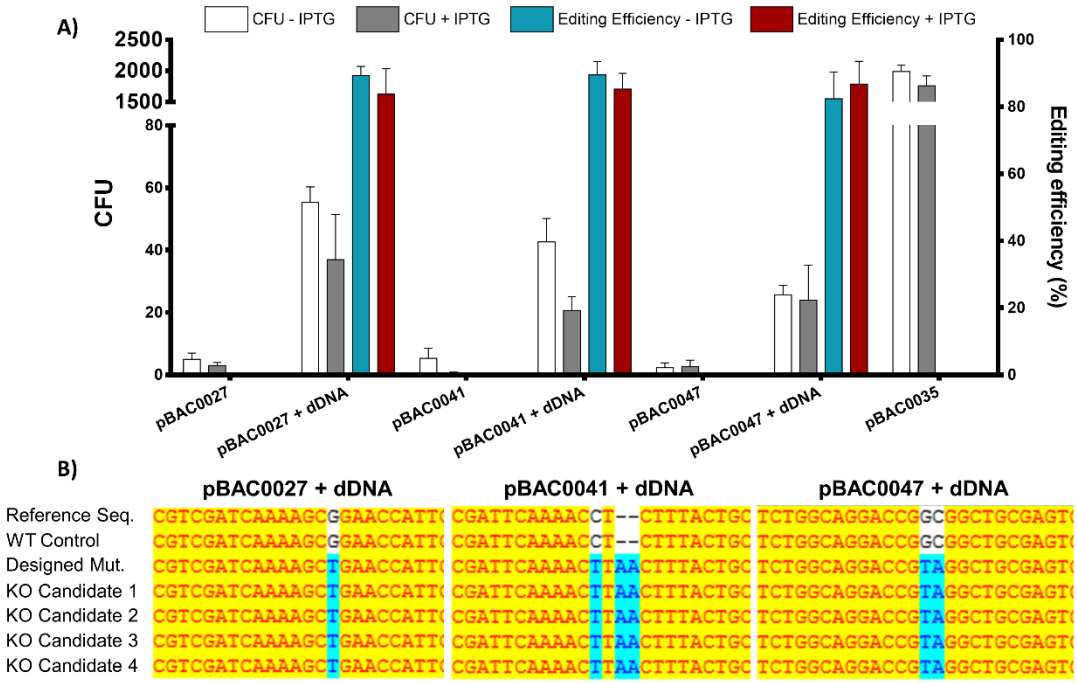


Figure 3.4 – Editing efficiency using three sgRNAs targeting *amyE*.

A) White (- IPTG) and grey (+IPTG) bars represent the number of CFU obtained following transformation of three *amyE* targeting plasmids (pBAC0027/0041/0047) with or without editing template (dDNA) to introduce stop codons and repair the sgRNA-targeted Cas9 DSB. A plasmid containing a non-targeting sgRNA (pBAC0035) was transformed to monitor transformation efficiency. The blue (- IPTG) and red (+ IPTG) bars represent the editing efficiency of the obtained CFU determined by observing the presence or absence of a halo following iodine staining of the starch-containing transformation plates. Error bars indicate the standard deviation between three transformations. **B)** DNA sequencing results for the target region of four colonies with knock-out (KO) phenotypes from each target plasmid + dDNA transformation. DNA for sequencing was isolated by PCR amplification of genomic DNA for each target. KO candidates are compared to a similarly prepared and sequenced wild-type (WT) control, as well as the *in-silico* reference sequence and the designed mutation.

3.3.2 Analysis of CRISPR-enabled genome editing efficiencies with alternative nucleases

Following the release of MAD7 for academic or industrial research and strain development by Inscripta, the editing efficiency of MAD7 within the industrial

workhorse *B. subtilis* was investigated. As such a side by side comparison of *S. pyogenes* Cas9, *Acidaminococcus sp.* BV3L6 Cpf1 (AsCpf1) and MAD7 was carried out to determine if MAD7 could act as a CRISPR nuclease for commercial strain development. AsCpf1 was also selected for this comparison as it has the greatest level of identity (31%) to MAD7[70]. The different characteristics of these nucleases are outlined in Table 3.2. *MAD7* was codon optimised (sequence comparison in Appendix D) for *B. subtilis* and synthesised (Thermo Fisher Scientific), while the readily available human codon optimised AsCpf1 from pY010 was utilised.

Table 3.2 – Different characteristics of the alternative CRISPR nucleases tested for activity in *B. subtilis*.

	***S. pyogenes* Cas9**	***Acidaminococcus sp.* BV3L6 Cpf1**	**MAD7**
Nuclease Type	Type II	Type V	
Cut type	Blunt	Staggered	
Native guide RNA	2 RNA molecules	1 RNA molecule	
PAM	5´ NRG 3' (R = G or A)	5´ TTTV 3' (V = G, C or A)	5´ YTTN 3' (Y = T or C)
Amino acid length	1368	1307	1263

'Base CRISPR-MAD7 and CRISPR-AsCpf1 plasmids were constructed in a similar manner to the 'base CRISPR-Cas9 plasmid, pBAC0008 (section 3.2.1.1). The primary difference between these plasmids is that the handle through which the Cas9 associates to the sgRNA was replaced with the specific gRNA handles for MAD7 and AsCpf1 respectively. As P_{veg} was shown to be well suited to sgRNA expression, this was retained within the editing plasmids and *aphAI*, conveying kanamycin resistance, was not introduced. Furthermore, two SapI restriction enzyme recognition sites were used to replace the AarI sites to facilitate protospacer cloning. While SapI yields 3 nt ssDNA overhangs, instead of the 4 nt following AarI digestion, SapI is known to be more reliable, cheaper and faster-acting than AarI.

MAD7 and *AsCpf1* were inserted into pBAC0155 and pBAC0154 respectively, yielding pBAC0158 and pBAC0159 (section 3.2.1.1).

To compare the genome editing efficiencies between Cas9, AsCpf1 and MAD7 in *B. subtilis*, the *amyE* and *gfpmut3* genes were selected for knock-out due to their ease of analysis to identify successfully edited colonies. Due to their similarity in PAM site specificities, AsCpf1 (5´ TTTV 3' (V = G,C or A)) and MAD7 (5´ YTTN 3' (Y = T or C)) targeted the same sites for both individual gene targets (*amyE* – TTTG; *gfpmut3* – TTTC). This site was selected based on its proximity to the Cas9 PAM site to allow as close a comparison as possible. A single dDNA was designed for each target to introduce a stop codon at the 5´ of the gene and remove the PAM site. Editing of *amyE* and *gfpmut3* was performed in *B. subtilis* 168 and BAC0288 respectively.

Plasmids carrying non-targeting gRNA were prepared for each nuclease to measure transformation efficiency and the potential toxicity towards *B. subtilis*. The transformation efficiency of the non-targeting AsCpf1 plasmid (1.3×10^2 CFU/µg DNA in *B. subtilis* 168 and 4.5×10^1 CFU/µg in BAC0288) was lower than that of Cas9 (1.3×10^4 CFU/µg DNA in *B. subtilis* 168 and 2.5×10^3 CFU/µg in BAC0288) and MAD7 (9.4×10^3 CFU/µg DNA in *B. subtilis* 168 and 2.4×10^3 CFU/µg in BAC0288). This suggests a toxicity effect of AsCpf1 towards *B. subtilis* which is not observed with Cas9 nor MAD7.

3.3.2.1 Comparison of the editing efficiencies obtained following editing of *amyE* with Cas9, AsCpf1 and MAD7

The pBAC0041 Cas9 editing plasmid (section 3.3.1.4) was utilised alongside plasmids pBAC0161 and pBAC0162 for AsCpf1 and MAD7 enabled genome editing of *amyE* respectively. A single dDNA was designed for each target to introduce a stop codon at the 5´ of the gene and remove the PAM sites (Figure 3.5A). Cas9, AsCpf1 and MAD7 yielded *amyE* knock-out efficiencies of 91%, 1.6% and 98%, respectively when *B. subtilis* 168 was co-transformed with the editing plasmids and dDNA (Figure 3.5B/C). Thus, in this case, MAD7 was found to have an editing efficiency of a least as good as Cas9. The 1.6% editing observed with AsCpf1, in combination with the low transformation efficiency obtained, was thought to be due to serendipitously dDNA being incorporated into the chromosome by homologous recombination, silencing *amyE*, and the same CFU incorporating the plasmid. No

successfully edited colonies were observed for any nuclease in the absence of dDNA.

To identify the benefit of having the targeting nucleases present in this system, the non-targeting Cas9 plasmid (pBAC0035) and the MAD7 equivalent (pBAC0163) were transformed alongside the dDNA, as well as separate co-transformations utilising the targeting plasmid (Figure 3.5D). AsCpf1 was not analysed since it had not exhibited a high efficiency of editing. The results indicate the clear advantage for editing efficiency in the presence of the targeting plasmid with efficiencies increasing from 0% to 98% for Cas9 and 0% to 93% for MAD7 (Figure 3.5D). Following phenotypic analysis, the genotypes of a selected population of the transformants were confirmed by colony PCR and sequencing. Primers used during colony PCR (oMAP0811/0812) hybridised outside of the homology arm region on the chromosome. All colonies screened confirmed the expected genotype following phenotypic analysis (Cas9 – WT: 2/2 colonies; knock-out: 9/9 colonies. MAD7 – WT: 2/2 colonies; knock-out: 10/10 colonies).

Following editing with MAD7, a higher average number of CFU was obtained than with Cas9. This is thought to represent a potential advantage over Cas9 genome editing, potentially due to its smaller size having less burden on the cell. However, following an unpaired t test with Welch's correction, it was established that this difference in CFU was not statistically significant.

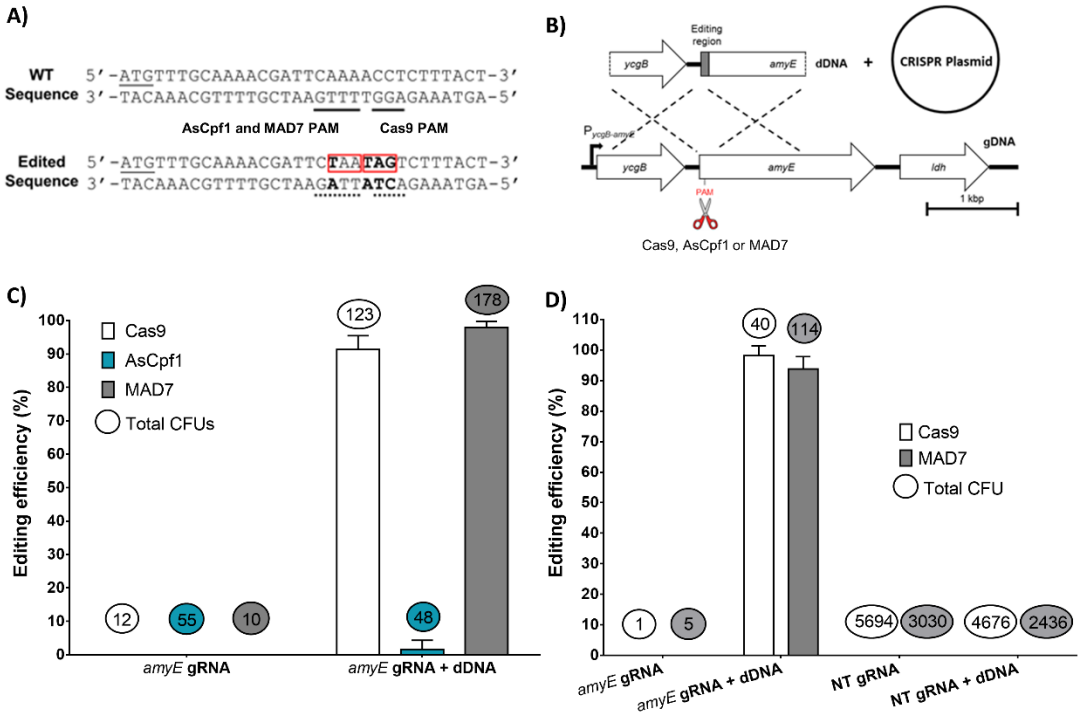


Figure 3.5 – Editing efficiencies obtained following CRISPR-Cas9/AsCpf1/MAD7-mediated *amyE* editing.

A) Non-edited (WT) and edited sequences. The ATG start codon is underlined. The targeted PAM sites are indicated for Cas9, AsCpf1 and MAD7. The modified base pairs are highlighted in bold and the introduced stop codons are marked with red boxes.
B) Co-transformational editing approach where the plasmid expressing the gRNA and nuclease are used to transform *B. subtilis* 168 alongside a linear editing template (dDNA) containing the editing region. **C)** Transformations and co-transformations comparing editing efficiencies when utilising Cas9, AsCpf1 or MAD7 nucleases in the presence or absence of editing template (dDNA). **D)** Transformations identifying the benefit in editing efficiency to having the nuclease and targeting gRNA present. Bars represent the average number of CFU obtained following transformation of the *amyE* targeting editing plasmid for each nuclease (Cas9 – pBAC0041; AsCpf1 – pBAC0161; MAD7 – pBAC0162), or non-targeting (NT) plasmids for Cas9 (pBAC0035) or MAD7 (pBAC0163), with or without dDNA to introduce a stop codon and remove the PAM site. The circled number above each bar represents the total number of transformants phenotypically screened by observing the presence or absence of a halo following iodine staining of the starch-containing transformation plates. Error bars indicate the standard deviation between three independent transformation events.

3.3.2.2 Comparison of the editing efficiencies obtained following editing of *gfpmut3* with Cas9, AsCpf1 and MAD7

As an additional target for genome editing efficiency, strain BAC0288 expressing the *gfpmut3* gene under the control of P_{veg} was prepared (section 3.2.1.2). pBAC0165, pBAC0167 and pBAC0166 were used to transform BAC0288 with and without dDNA to test Cas9, AsCpf1 and MAD7 editing efficiencies respectively through the introduction of stop codons and synonymous PAM site mutations to prevent further cutting by each nuclease (Figure 3.6A/B). A single dDNA was used for all nucleases. When knocking out *gfpmut3*, editing efficiencies of 83%, 0% and 100% were observed in strain BAC0288 for Cas9, AsCpf1 and MAD7, respectively (Figure 3.6C). No successfully edited colonies were observed for any nuclease in the absence of dDNA. As for *amyE*, the base level editing rate for Cas9 and MAD7 targeting modification of *gfpmut3* was investigated using non-targeting plasmids and the same dDNA (Figure 3.6D). The results showed a similar benefit to editing efficiency when the targeting plasmid was utilised with rates increasing from 0.18% to 75% for Cas9 and 0.23% to 100% for MAD7 (Figure 3.6D). Following phenotypic analysis, the genotypes of all colonies exhibiting the knock-out phenotype, from the targeting and non-targeting CRISPR plasmid co-transformations with dDNA, and a selected population of the transformants exhibiting the WT phenotype were confirmed by colony PCR and sequencing. Primers used during colony PCR (oMAP0393/0814) hybridised outside of the homology arm region on the chromosome. All colonies screened confirmed the expected genotype following phenotypic analysis (Cas9 – WT: 7/7 colonies; knock-out: 5/5 colonies. MAD7 – WT: 6/6 colonies; knock-out: 4/4 colonies).

BAC0288 proved to be considerably less competent than *B. subtilis* 168 with a transformation efficiency of 5.2-fold less for the Cas9 control plasmid (pBAC0035) and 3.9-fold less for the MAD7 control plasmid (pBAC0163). As a result, while the capacity to edit the target site efficiently is apparent, these results on their own are not statistically robust due to the low CFU obtained and should be taken as supporting evidence with the efficiencies obtained in the *amyE* editing experiment in section 3.3.2.1.

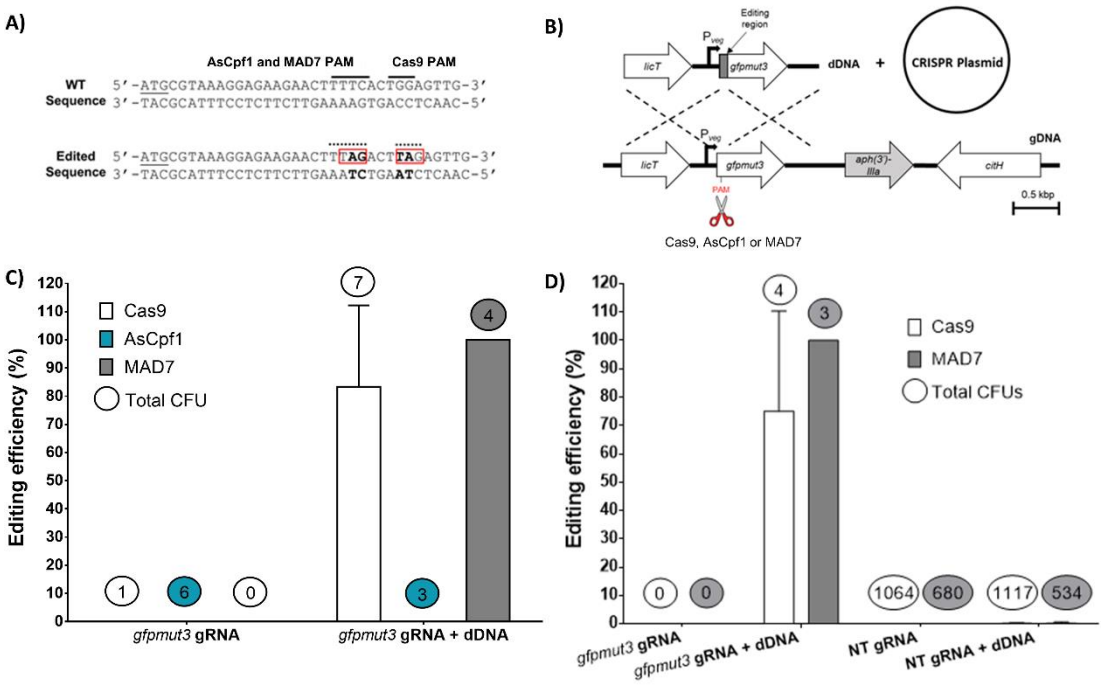


Figure 3.6 – Editing efficiencies obtained following CRISPR-Cas9/AsCpf1/MAD7-mediated *gfpmut3* editing.

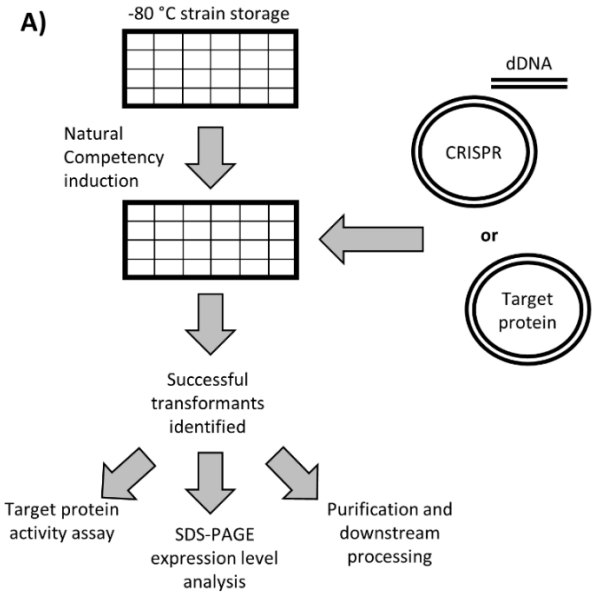
A) Non-edited (WT) and edited sequences. The ATG start codon is underlined. The targeted PAM sites are indicated for Cas9, AsCpf1 and MAD7. The modified base pairs are highlighted in bold and the introduced stop codons are marked with red boxes.
B) Co-transformational editing approach where the plasmid expressing the gRNA and nuclease are used to transform *B. subtilis* 168 alongside a linear editing template (dDNA) containing the editing region. **C)** Transformations and co-transformations comparing editing efficiencies when utilising Cas9, AsCpf1 or MAD7 nucleases in the presence or absence of dDNA. **D)** Transformations identifying the benefit in editing efficiency to having the nuclease and targeting gRNA present. Bars represent the average number of CFU obtained following transformation of the *gfpmut3* targeting editing plasmid for each nuclease (Cas9 – pBAC0165; AsCpf1 – pBAC0167; MAD7 – pBAC0166), or non-targeting (NT) plasmids for Cas9 (pBAC0035) or MAD7 (pBAC0163), with or without dDNA to introduce a stop codon and remove the PAM site. The circled number above each bar represents the total number of transformants phenotypically screened by observing the presence or absence of GFPmut3 following blue light analysis. Error bars indicate the standard deviation between three independent transformation events.

3.3.2.3 Development of a protease knock-out strain library with MAD7

As it had been established that MAD7 was as or more efficient than Cas9 for genome editing in *B. subtilis* 168, an industrially relevant use for this new tool was sought. The extracellular protease deficient BRB strains, developed by Pohl *et al*., underwent sequential whole gene deletion of the seven extracellular proteases (NprB, AprE, Epr, Bpr, NprE, Mpr and Vpr) present in *B. subtilis* 168 cultures[40]. Further deletions were subsequently made of the membrane bound proteases WprA, HtrA and HtrB which are involved in quality control of proteins secreted into the extracellular environment[40,144,145]. These strains, or similar alternative strains, have since been used for heterologous protein production, and recently also in stably tethering proteins to the cell wall to enable future uses in biocatalysis, bioremediation or lignocellulose degradation[40,41,146].

It was decided to produce a MAD7 developed version of these strains were frame shift mutations for each protease creates effective knock-outs with final strains free for commercial use and with minimal genomic modifications utilised instead of the full gene deletions employed for the construction of the BRB strains. Industrially, these new strains could be utilised in a 24-well plate format to screen the strains with increasing number of silenced genes with any extracellular heterologous protein target to identify which strain produces the highest amount of the target protein (Figure 3.7). Alternatively, each strain could be further modified using CRISPR tools to investigate the effect of further gene knock-outs, or knock-ins.

Sequential deletions such as those defined here can be rapidly carried out with CRISPR tools. Once all genetic elements are available, two rounds of editing can be carried out in five days. This includes quality control checks by colony PCR and sequencing to ensure the mutation was introduced as designed, as well as subsequent plasmid loss before the next round of mutations. This is possible due to high efficiency of the CRISPR systems making it rare to select a non-edited colony, and as such plasmid loss and quality control checks can be ran in parallel. Furthermore, if there had been unforeseen delays in the sequencing, the use of the editing plasmid from the previous round allows a clear report that the synonymous PAM mutation had been introduced, and as such there is a high likelihood that the nearby frameshift mutation was also introduced.



B)

24-well position	Genotype
A1	trpC2
A2	trpC2, ΔaprE
A3	trpC2, ΔnprE,
A4	trpC2 Δvpr,
A5	trpC2 Δbpr,
A6	trpC2 ΔnprB,
B1	trpC2 Δmpr,
B2	trpC2 Δepr,
B3	trpC2 ΔaprE, ΔnprE
B4	trpC2 ΔaprE, ΔnprE, Δvpr
B5	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr
B6	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB
C1	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB, Δmpr
C2	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB, Δmpr, Δepr
C3	trpC2 ΔhtrA
C4	trpC2 ΔhtrB
C5	trpC2 ΔwprA
C6	trpC2 ΔhtrA, ΔhtrB
D1	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB, Δmpr, Δepr, ΔhtrA
D2	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB, Δmpr, Δepr, ΔhtrB
D3	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB, Δmpr, Δepr, ΔwprA
D4	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB, Δmpr, Δepr, ΔhtrA, ΔhtrB
D5	trpC2 ΔaprE, ΔnprE, Δvpr, Δbpr, ΔnprB, Δmpr, Δepr, ΔhtrA, ΔhtrB, ΔwprA
D6	Contamination control well (media blank)

Figure 3.7 – Protease knock-out strain library.

A) The proposed workflow for new heterologous protein expression targets in *B. subtilis*. Target proteins for secretion can be introduced either on replicative plasmids, or incorporated onto the chromosome using the CRISPR approach described here. Downstream assays for identification of the highest producing strain is dependent on the target protein for production. **B)** The intended strain genotype for each well.

The strain library construction was initiated with plasmids pBAC0197-0206 being combined with their relevant dDNA (section 3.2.1.2) and co-transforming naturally competent *B. subtilis* 168. Transformants were selected on LB agar supplemented with chloramphenicol and IPTG. When only the targeting plasmid was transformed, a drastic drop in transformation efficiency was observed for six of the ten targets (Table 3.3). These genes (*aprE*, *nprE*, *htrA*, *bpr*, *mpr*, and *epr*) were found to have been knocked out as planned following colony PCR and sequencing. No successful mutants were obtained for the remaining targets (*htrB*, *nprB*, *wprA* and *vpr*).

It was hypothesised that the remaining targets had not worked as planned due to issues either in the PAM site targeted (*nprB* and *wprA*) or a potential new

PAM site introduced following HR of the dDNA (*vpr* and *htrB*). Data provided by Inscripta indicated that TTTT PAM sites are generally functional (as shown here with *bpr* and *vpr* targets) though not in all cases (potentially observed here with *nprB* and *wprA*)[67]. Future targeting of these genes will utilise the same dDNA, but the target PAM will be moved by a single base to ensure a TTTV (V = A, G or C) PAM site.

No knock-out colonies were identified for *vpr* or *htrB* targets. As in both cases the PAM site was altered to TCTA, and other bases within the YTTN (Y = C or T) PAM site can tolerate C bases, it is hypothesised that this represents a potential additional PAM site[67]. However, due to time constraints and following discussions with my supervisor, this hypothesis was not further investigated.

This strain library (Figure 3.7) could not be completed within the time limitations of this study. However, once completed and successfully used to identify the optimum background strain for a target heterologous protein, subsequent rounds of editing would be desirable to produce a final strain for use in large-scale fermentation processes. The repair of the *trpC2* gene to return the strain to a prototrophic state, as described in section 3.3.3, is desirable to allow the use of cheaper chemically defined media without the need for tryptophan supplementation. Next, to prevent the strain entering sporulation and maximise protein production, the sporulation master regulator, *spo0A* should be targeted for knock-out[147]. Alternatively, sigma factor F (*spoIIAC*) is also known to be responsible for transcription of forespore proteins and as such is an alternative/additional target for deletion[41]. Deletion of other highly expressed secreted proteins, such as α-Amylase (*amyE*) could aid downstream protein purification, as well as potentially increasing the rate with which proteins are expressed and secreted by not blocking the relevant cellular machinery. Lastly, to avoid excessive foaming in large-scale fermentation processes, thereby lowering costs and reducing potential loss of culture, the surfactin synthase subunit (*srfAC*) should be deleted[41].

Table 3.3 – Colonies obtained following the first round of protease knock-out library construction.

Target gene	DNA Added	CFUs obtained	PAM site (5´-3')
aprE	pBAC0197	2	TTTG
	pBAC0197 + dDNA	97	
nprE	pBAC0198	3	TTTA
	pBAC0198 + dDNA	106	
htrA	pBAC0199	2	TTTC
	pBAC0199 + dDNA	74	
htrB	pBAC0200	1	TTTG
	pBAC0200 + dDNA	14	
bpr	pBAC0201	6	TTTT
	pBAC0201 + dDNA	127	
nprB	pBAC0202	1432	TTTT
	pBAC0202 + dDNA	2128	
mpr	pBAC0203	6	TTTG
	pBAC0203 + dDNA	118	
epr	pBAC0204	4	TTTC
	pBAC0204 + dDNA	221	
wprA	pBAC0205	832	TTTT
	pBAC0205 + dDNA	800	
vpr	pBAC0206	3	TTTT
	pBAC0206 + dDNA	5	
No target	pBAC0163	4032	N/A

3.3.3 Homologous recombination vs. DNA double-strand break repair as the driving mechanism for CRISPR genome editing in *B. subtilis* 168

To elucidate the mechanism with which CRISPR-mediated editing takes place in *B. subtilis* 168, we made use of this strain's tryptophan auxotrophy to compare the efficiency in restoring prototrophy when a linear dDNA is transformed on its own, or in the presence of either a Cas9 or MAD7 non-targeting or *trpC2*-targeting plasmid. The linear dDNA was designed to simultaneously introduce an additional isoleucine residue adjacent to I110 residue of TrpC2, returning the strain to a prototrophic state, and a synonymous mutation to eliminate the PAM recognition site and prevent continuous cutting by the *trpC2*-targeting plasmid (Figure 3.8A/B)[63]. By selecting transformants in M9 minimal medium supplemented with or without either chloramphenicol (plasmid selection) or tryptophan, we could clarify whether HR drives genome editing, preventing a DSB, or if the DSB induces DNA repair by HR.

In the absence of tryptophan there is not a significant difference in CFU obtained when transforming the linear dDNA to restore tryptophan prototrophy on its own or in the presence of either the *trpC2* targeting or non-targeting plasmids (Figure 3.8). Furthermore, when cells with restored prototrophy were also selected in the presence of chloramphenicol, there was no significant difference between the co-transformation of dDNA with either the *trpC2*-targeting or non-targeting CRISPR-Cas9 plasmids (Figure 3.8). Both these results indicate that HR is the main driving force for CRISPR-Cas9 editing in the presence of dDNA. When tryptophan was supplemented to M9, the absence of a selective pressure for restored prototrophy results in a significantly lower number of CFU when co-transforming the dDNA and the *trpC2*-targeting plasmid compared to the co-transformation of dDNA with the non-targeting plasmid (Figure 3.8). In this case, the lethal cut induced by the nuclease counterselects the transformants in which HR of dDNA has not occurred. As such, while the high efficiency of HR is the main driving force for genome editing, the nuclease induced DSB is essential to obtain high editing efficiency in *B. subtilis* 168.

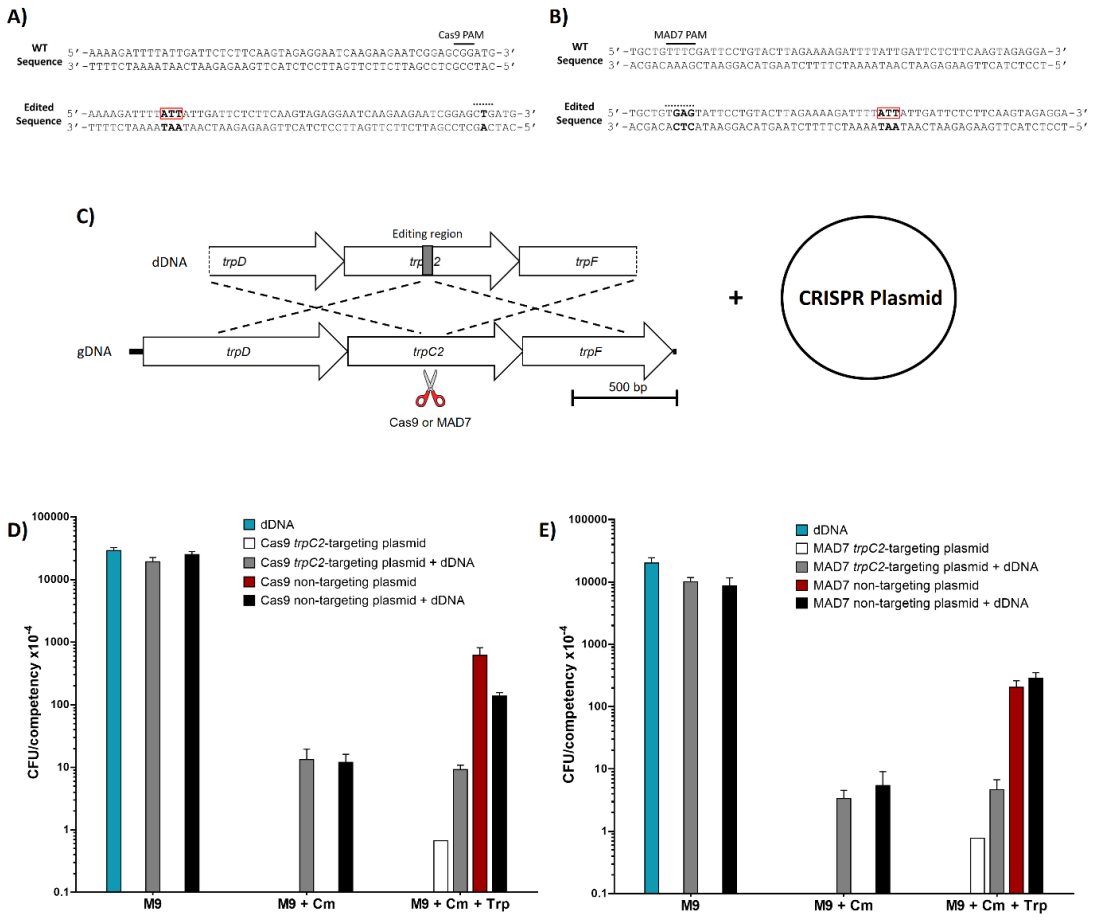


Figure 3.8 – Restoration of *B. subtilis* prototrophy using CRISPR-Cas9 and CRISPR-MAD7 for genome editing.

A) and **B)** show the non-edited (WT) and edited sequences for Cas9 and MAD7 editing respectively. The modified base pairs are highlighted in bold and the inserted isoleucine codon, adjacent to I110, is marked with red boxes. **C)** Co-transformational editing approach where the CRISPR plasmid expressing the gRNA and nuclease is transformed alongside a linear editing template (dDNA) containing the editing region. **D)** and **E)** Graphs show the number of transformants following transformations with the indicated combinations of dDNA and targeting (pBAC0185 for Cas9 and pBAC0218 for MAD7) or non-targeting plasmid (pBAC0035 for Cas9 and pBAC0163 for MAD7) to restore *B. subtilis* 168 prototrophy, with Cas9 and MAD7 respectively. Bars represent the average number of colony forming units (CFU) normalised by the transformation efficiency of pBAC0035 (Cas9) selected on LB agar supplemented with chloramphenicol or pBAC0163 (MAD7) selected on LB agar supplemented with chloramphenicol and IPTG. Error bars indicate the standard deviation between three independent transformation events. Cm – chloramphenicol; Trp – Tryptophan.

3.4 Discussion

3.4.1 Limitations and inefficiencies within the CRISPR-Cas genome editing approach

In this chapter, CRISPR systems were developed and characterised for use with *B. subtilis* 168. This approach, where a single plasmid for *cas9* and sgRNA expression, and a linear OE-PCR product as dDNA are co-transformed, relies on the preparation of highly competent *B. subtilis* cells. Provided the promoter and origin or replication genetic elements are active, the transformation efficiency is the key limitation with this approach being utilised in other Bacilli or strains of *B. subtilis* where a drastic drop in natural competence efficiency has occurred. In such cases, it may be best to introduce the dDNA into the plasmid to remove the need for two transformation events to take place, as exemplified by Altenbuchner[63]. However, this was not investigated in this study.

Editing in the presence of higher colony numbers did not reach 100% since a low number of CFUs were obtained when only the editing plasmid is present. These would appear to be instances when the colony has escaped cutting. Such 'escaper' colonies could have formed due to random mutations interrupting the PAM site. However, as the spontaneous mutation frequency for *B. subtilis* is between 10^{-8} and 10^{-7}, and, by using the pBAC0035 control plasmid CFUs, an average of 1763 cells with IPTG, and 1989 cells without IPTG were transformed with the editing plasmid, it appears unlikely that PAM sites were interrupted by random mutagenesis in the 'escaper' colonies[8]. Given that there is a strong selection for cell survival in the presence of the plasmid due to the plates being supplemented with chloramphenicol, it is more likely that these transformants are instances where the plasmid incorporated has a mutation inactivating either the nuclease or gRNA thus preventing cleavage and allowing colony formation.

We have provided evidence that, in *B. subtilis* 168, CRISPR-Cas9 and CRISPR-MAD7 genome editing efficiency is driven primarily by HR of dDNA preventing the lethal Cas9 or MAD7-induced DNA DSB, rather than HDR following the DSB. Mougiakos *et al*. reported the endogenous HR machinery within *Bacillus smithii* incorporating plasmid borne dDNA while Cas9 was inactive at ≥42 °C. Counterselection of the cells which had not undergone HR was then performed at 37 °C where the Cas9 was once again functional. Here we have shown HR as the driving mechanism under temperatures where the nuclease is

active and growth is optimal[117]. The presence of the CRISPR-Cas9 or CRISPR-MAD7 system was not required to detect successful HR due to prototrophic selection on M9 minimal medium. However, where such a selection is not possible, the CRISPR-Cas9 or CRISPR-MAD7 systems act as a powerful counterselection for unedited cells. The lack of a significant increase in editing efficiency following addition of IPTG to drive *cas9* expression, and as sgRNA expression is known to be under the strong constitutive promoter, P_{veg}, indicated editing rates are decoupled from Cas9 expression levels with leaky P_{grac} promoter activity being sufficient to induce the lethal DSB. Moreover, as the natural competency master regulator ComK activates transcription of the primary component of HR, *recA*, and DNA uptake is single stranded, yielding a substrate with which RecA can bind, HR can readily proceed at the target site prior to the DSB taking place[148,149].

3.4.2 Toxicity of AsCpf1 to *Bacillus subtilis*

It was not clear why AsCpf1 appeared to have a degree of toxicity to *B. subtilis* following transformations utilising non-targeting plasmids and transformation efficiencies ~100 fold less than MAD7 or Cas9 equivalent transformations. Further experiments to elucidate this were not carried out as the primary goal was to compare the new, commercially free-to-use MAD7 with the well characterised and established Cas9. As a hypothesis however, it is possible that the toxicity is linked to *AsCpf1* being codon optimised for humans in pY010[68]. As such, the codon usage ratio (frequency per thousand codons of *B. subtilis*/AsCpf1) was analysed and showed the human codon optimised *AsCpf1* utilises 800/1307 codons used at least 50% less often in *B. subtilis* (Figure 3.9). Considering the high number of rarely used codons, it may be the case that the mRNA translation is being continually paused, causing an accumulation of misfolded or incomplete insoluble protein, which is known to be toxic to the cell[150].

Future analysis of genome editing with AsCpf1 should synthesise a *B. subtilis* codon optimised gene to enable clear comparison with other nucleases. Furthermore, comparison assays with the catalytically inactive dAsCpf1 variant would elucidate if in fact a random catalytic activity of AsCpf1 is the toxic element observed here. The developed parental plasmid for use with AsCpf1, pBAC0154, can be utilised for these future experiments.

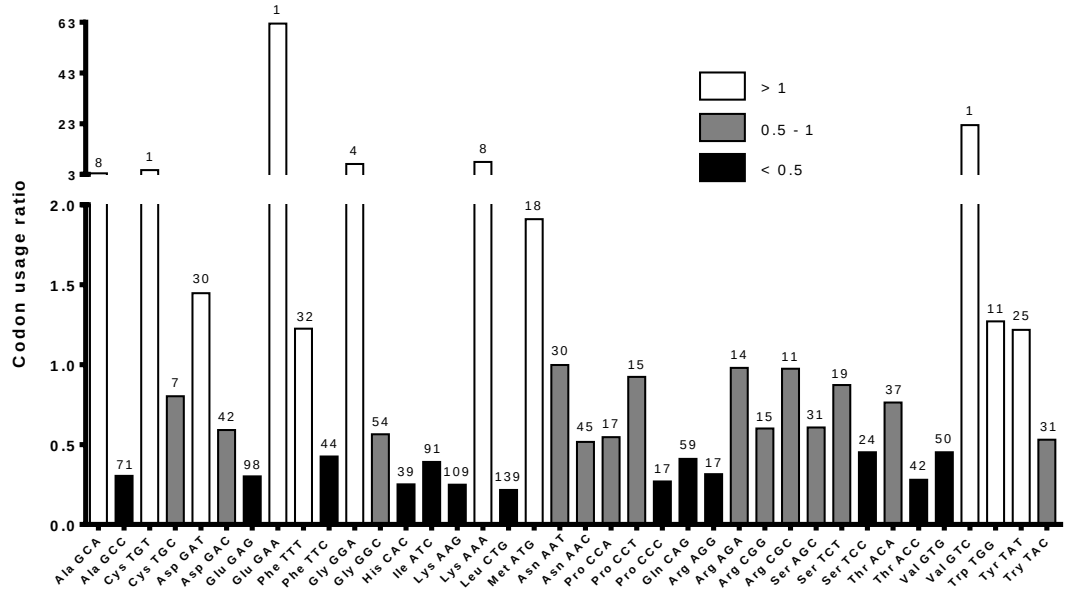


Figure 3.9 – Ratio of codon usage in *B. subtilis* to *AsCpf1* codon optimised for humans.

White bars show all used codons with ratio >1 and unlikely to be limited for expression in *B. subtilis*. Similarly, grey bars represent a ratio of 0.5 – 1 where codons are more limited, and black bars represent a ratio < 0.5 where codons could be severely limited. The numbers above each bar indicate the number of amino acid residues using the corresponding codon.

3.5 Conclusions

CRISPR-Cas9 mediated genome editing of *B. subtilis* 168 was found to be a highly efficient process at 87.1% in the absence of IPTG and 85.2% in the presence of IPTG. Such high editing efficiencies make it feasible to proceed with strain development in a rapid manner with several rounds of editing. Similar high editing rates were observed with the alternative CRISPR nuclease, MAD7. Since *B. subtilis* is used widely as an industrial workhorse, the use of MAD7 may be more attractive than Cas9 for commercial strain construction. However, significant questions regarding application of MAD7 remain. A significant case study, and potentially a challenge to the MAD7 IP, owned by Inscripta, may be required before industry fully trusts the claims made by Inscripta that MAD7 is free to use for commercial R&D and strain construction.

Chapter 4 – CRISPR-Cas9 *In Situ* engineering of subtilisin E in *Bacillus subtilis*

4.1 Introduction

As a proof that CRISPR-Cas9 was a valuable tool for rapid genome editing for the development of an industrially significant protein production target, the extracellular protease Subtilisin E (encoded by the *aprE* gene) was selected as it is used globally within the detergent, food, pharmaceutical and leather industries[151]. Subtilisin E serves in nature to degrade other proteins utilising a nucleophilic S327 residue in its catalytic triad (D138, H170, S327) for hydrolysis of the target peptide bond[39,152].

As the wild type Subtilisin E enzyme can be broken down by detergent formulations and heat, subtilisin variants with improved thermostability and pH tolerance have long been of interest. Subtilisin E has been widely used as a target for protein engineering experiments which can be split into 8 categories for designed improvements: Catalytic mechanism, substrate specificity, new activities, general proteolytic activity general stability, stability in exotic environments, surface activity and folding mechanisms. These are reviewed in detail by Bryan[39].

A salt-bridge triad (Arg19-Glu271-Arg275), identified in the subtilisin E homolog from *Bacillus clausii* (M-protease), was found to be a contributor for the characteristic thermotolerance of this enzyme through site-directed mutagenesis experiments[153]. Mutation of these sites (R19Q-E271Q-R275Q) allowed a similar level of activity retention following incubation at 55 °C, while ≥60 °C caused an increased rate of activity reduction when compared the WT enzyme. Following incubation at 60 °C, a 50% drop in activity was observed following 12.9 minutes incubation, whereas the WT enzyme required 37.8 minutes incubation for the same reduction in activity[153].

In this chapter, we use CRISPR-Cas9 mediated genome editing to replace the respective residues in subtilisin E (Gln125-Gln377-Gln381), and evaluate the effect of the salt bridge on the thermostability and activity of the new variant and two other intermediate variants. This work illustrates the use of a simple CRISPR-Cas9

system for *B. subtilis* rapid, *in situ* protein engineering, which is at the core of industrial biotechnology to provide new, suitable and competitive biocatalysts.

The work in this chapter was published in January 2019 and a copy of this publication can be found in Appendix E.

4.2 Materials and methods

4.2.1 Plasmids and strains

The plasmids and strains used in this chapter are outlined below in Table 4.1. The oligonucleotides used in this chapter can be found in Appendix C.

Table 4.1 – Strains and plasmids used in this chapter.

Strain/ Plasmid	**Genotype/Description**	**Reference**
Strains		
B. subtilis 168	*trpC2*	Laboratory stock
BAC0094	*B. subtilis* 168 with *aprE* Q125R variant.	This chapter.
BAC0095	*B. subtilis* 168 with *aprE* Q377E and Q381R variant.	This chapter.
BAC0097	BAC0095 with *aprE* Q125R, Q377E and Q381R variant	This chapter.
BAC0114	*B. subtilis* 168 *ΔaprE::aad9*	This chapter.
BAC0116	BAC0114 with pHT01.	This chapter.
BAC0117	BAC0114 with pBAC0057.	This chapter.
BAC0118	BAC0114 with pBAC0058.	This chapter.

BAC0119	BAC0114 with pBAC0059.	This chapter.
BAC0120	BAC0114 with pBAC0060.	This chapter.
BAC0121	BAC0114 with pBAC0068.	This chapter.
BAC0122	BAC0114 with pBAC0069.	This chapter.
Plasmids		
pHT01	*E. coli/B. subtilis* shuttle vector carrying P_{grac} and *lacI*, *bla*, *cat*.	MoBiTec
pDR111	*bla*; 5´ *amyE*; *aad9*; P_{spac};*lacI*; 3' *amyE*	[154]
pBAC0015	pBAC0008 with *cas9* (*S. pyogenes*). Cas9 expression regulated by the P_{grac}.	Section 3.2.1.1
pBAC0054	pBAC0015 with sgRNA targeting 359 bp downstream of the start codon within *B. subtilis* 168 *aprE*.	This chapter.
pBAC0055	pBAC0015 with sgRNA targeting 1163 bp downstream of the start codon within *B. subtilis* 168 *aprE*.	This chapter.
pBAC0057	pHT01 with *aprE* (native) from *B. subtilis* 168 with native GTG start codon.	This chapter.
pBAC0058	pHT01 with *aprE* (Q125R, Q377E and Q381R variant) from BAC0097 with native GTG start codon.	This chapter.
pBAC0059	pHT01 with *aprE* (native) from *B. subtilis* 168 with ATG start codon.	This chapter.
pBAC0060	pHT01 with *aprE* (Q125R, Q377E and Q381R variant) from BAC0097 with ATG start codon.	This chapter.

pBAC0068	pHT01 with *aprE* (Q125R variant) from BAC0094 with ATG start codon.	This chapter.
pBAC0069	pHT01 with *aprE* (Q377E and Q381R variant) from BAC0095 with ATG start codon.	This chapter.

4.2.2 Plasmid construction

Plasmids used for Cas9 -mediated editing were prepared from the pBAC0015 parental plasmid, using phosphorylated and annealed oligonucleotide pairs as described in Section 2.5.4.1. Information on the oligonucleotides used in construction, target gene and PAM site for the following plasmids used in this chapter can be found in Table 2.12: pBAC0054; pBAC0055.

Plasmids pBAC0057/0058/0059/0060/0068/0069 were constructed by PCR amplifying the variant of *aprE* from the gDNA of the strain specified in Table 4.1. The PCR primers used for each variant amplification and introduction of BamHI and XmaI restriction enzyme sites, were: pBAC0057 – oMAP0199/0186; pBAC0058 – oMAP0199/0200; pBAC0059 – oMAP0203/0186; pBAC0060 – oMAP0203/0200; pBAC0068 – oMAP0203/0186; pBAC0069 – oMAP0203/0200. Each PCR product underwent BamHI-HF and XmaI restriction enzyme digestion (Section 2.5.8) and ligation (Section 2.5.10) with the similarly digested pHT01 vector.

4.2.3 Strain construction

Using the approach described in Section 2.7, CRISPR-Cas9 mediated editing of the *aprE* gene in *B. subtilis* 168 was carried out to construct strains BAC0094 and BAC0095. dDNA OE-PCR products, amplified from *B. subtilis* 168 gDNA with oligonucleotide sets oMAP0152/0153/0154/0155 and oMAP0158/0159/0160/0161 were used alongside pBAC0054 and pBAC0055 respectively. Editing was confirmed by PCR of the *aprE* gene with oligonucleotides oMAP0152/0161, followed by sequencing using oligonucleotides oMAP0158 and oMAP0155 for candidates edited with pBAC0054 and pBAC0055 respectively. The CRISPR-Cas9/sgRNA plasmid was removed from the edited strain as described in Section 2.7.6. Following the second round of editing, these mutations and those

introduced previously were confirmed by sequencing (as above) of the newly constructed strain, BAC0097.

BAC0114 was constructed by transformation of *B. subtilis* 168 with OE-PCR product containing a spectinomycin resistance cassette (PCR amplified from pDR111 with primers oMAP0221/0222) flanked by homology arms upstream and downstream of *aprE* (PCR amplified from *B. subtilis* 168 gDNA with oligonucleotides oMAP0217/0218 and oMAP0219/0220). Confirmation of *aprE* deletion was obtained by purification of the genomic DNA for BAC0114 and PCR of *ΔaprE::aad9* locus with oligonucleotides oMAP0670 (hybridising to *aad9*) and oMAP0671 (hybridising to the genome, upstream of the homology arm region). Additionally, PCR with oligonucleotide pair oMAP0217/0220 (hybridising to the extremities of the homology arm region) revealed the expected increase in product size for *ΔaprE::aad9* relative to the WT PCR product.

BAC0114 was transformed with plasmids pBAC0057/0058/0059/0060/0068/0069, resulting in strains BAC0017/0018/0119/0120/0121/0122 respectively.

4.3 Results

4.3.1 Identification of target residues for modification in subtilisin E

The crystal structures for *B. subtilis* subtilisin E (PDB ID 1SCJ) and its *B. clausii* homolog, M-protease (PDB ID 1WSD), were overlaid using Swiss-Pdb viewer (Figure 4.1)[129–131]. The residues corresponding to the salt-bridge triad (R19-E271-R275) that have previously been shown to contribute towards M-protease thermotolerance were identified in the subtilisin E structure (Q125-Q377-Q381)[153].

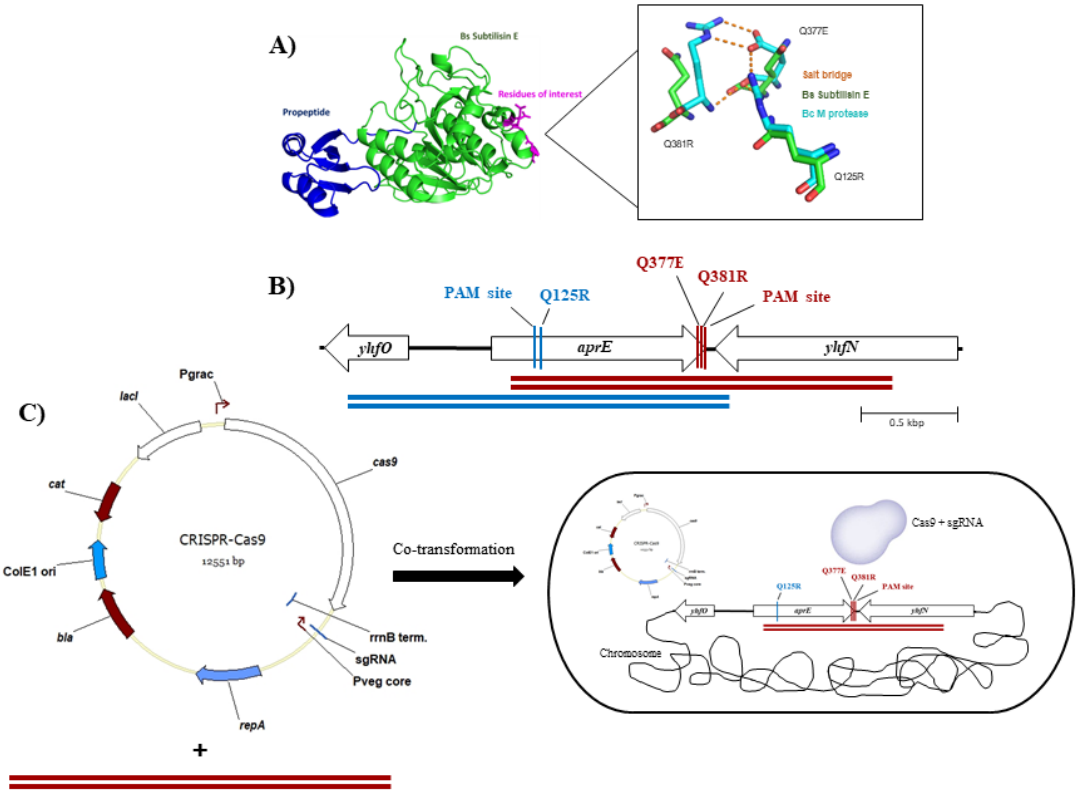


Figure 4.1 - Design and CRISPR-Cas9 editing process of *aprE*.

A) *B. subtilis* 168 subtilisin E crystal structure (green) and overlay with *B. clausii* M protease crystal structure (cyan), with the associated salt bridge (dashed yellow line). **B)** Design of the sgRNA and dDNA OE-PCR product for CRISPR-Cas9 genome editing. The two rounds of editing are described in blue and red, the PAM recognition sequences for each sgRNA was also targeted for disruption in each dDNA. **C)** Description of editing process where the second round of editing occurs following a blunt DSB by sgRNA guided Cas9.

4.3.2 *In situ* modification of *aprE* with CRISPR-Cas9

A CRISPR-Cas9 mediated chromosome editing strategy was designed to introduce the mutations Q125R, Q377E and Q381R (Figure 4.1). pBAC0015 was altered to include the oligonucleotide pairs oMAP0150/oMAP0151 or oMAP0156/oMAP0157, yielding plasmids pBAC0054 (targeting Q125R locus) and pBAC0055 (targeting Q377E and Q381R locus) respectively. Due to their proximity, Q377E and Q381R modifications were combined into a single CRISPR-Cas9 mediated editing step with a single dDNA. Q125R was modified in a separate editing

step. The desired edits, included the synonymous PAM mutation to prevent continuous cutting by the RNA-guided Cas9 endonuclease, were introduced in the homology overlap of the two DNA fragments prepared using PCR. These were subsequently combined using OE-PCR and co-transformed alongside the respective plasmid (Figure 4.1). OE-PCR products from oligonucleotides sets oMAP0152/0153/0154/0155 and oMAP0158/0159/0160/0161 were used alongside pBAC0054 and pBAC0055 respectively. Following an efficient curing process in which the editing plasmid was lost from the successfully edited strain by growth in the absence of chloramphenicol for plasmid selection, and in the presence of IPTG to increase the burden of Cas9 expression on cells retaining the plasmid, yielding strains BAC0094 (Q125R) and BAC0095 (Q377E and Q381R) respectively, a second round of editing using the second set of editing plasmid and dDNA yielded the final strain (BAC0097) containing all three modified residues. All screened colonies were found to contain the desired mutations following sequencing.

This system represents a rapid technique for *in situ* protein modifications within *B. subtilis* 168. Once the Cas9-sgRNA expression plasmid is prepared, the same region on the unmodified chromosome can be targeted with a library of alternative, rapidly prepared, linear dDNA templates conveying novel modifications of interest. As such it is feasible to target a region of a protein of interest, such as a substrate binding pocket, and introduce a library of modifications to identify beneficial variants of the target protein.

4.3.3 Purification of subtilisin E variants

4.3.3.1 Variant expression plasmid and strain preparation

The CRISPR-Cas9 method had been shown to allow rapid, *in situ*, protein engineering, applicable to target strains already in commercial production. However, the native subtilisin E is expressed at low levels, even when expression is induced on the multicopy plasmid, pHT01 in strain BAC0117 (Figure 4.2). To ensure accurate protein characterisation could be performed on the variants obtained in strains BAC0094, BAC0095 and BAC0097, each variant was cloned into the pHT01 vector, yielding plasmids pBAC0068, pBAC0069 and pBAC0060 respectively, allowing IPTG inducible expression and secretion of each variant. The native *aprE* gene was also cloned into pHT01, yielding plasmid pBAC0059, to act as a control throughout. During the cloning process, the native GTG start codon was replaced

with an ATG start codon to maximise protein expression levels. In *B. subtilis* ATG, TTG and GTG start codons are used in 78%, 13% and 9% of coding sequences respectively[155]. Furthermore, it has previously been shown that GTG start codons are three to five-fold less efficient than ATG start codons for translational initiation in *B. subtilis*[156]. In each newly constructed plasmid, each variant is expressed under control of the IPTG inducible P_{grac} promoter.

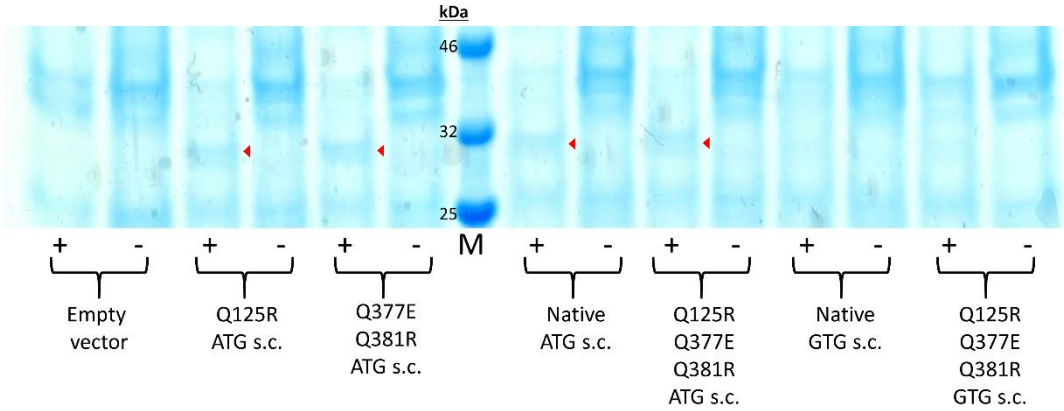


Figure 4.2 – Subtilisin E variant expression analysis.

Comparison of the spent media of strains expressing native and variant subtilisin E proteins to an empty vector control strain. All plasmids were introduced into the BAC0114 (*B. subtilis* 168 *ΔaprE::aad9*) strain. Subtilisin E (27.7 kDa) position on the gel is indicated by red triangles. All samples are derived from the clarified supernatant of the strains grown in presence (+) or absence (-) of IPTG for subtilisin E expression. For each variant, the subtilisin E start codon (s.c.) utilised is indicated. Strains used: Empty vector = BAC0116; Q125R ATG s.c. = BAC0121; Q377E Q381R ATG s.c. = BAC0122; Native ATG s.c. = BAC0119; Q125R Q377E Q381R ATG s.c. = BAC0120; Native GTG s.c. = BAC0117; Q125R Q377E Q381R GTG s.c. = BAC0118. M = Blue Prestained Protein Standard, Broad Range (11-190 kDa) (New England Biolabs).

These plasmids were used to transform strain BAC0114 which had undergone a full deletion of the wild-type *aprE* gene through introduction of the spectinomycin resistance gene, *aad9* (section 4.2.3). BAC0114 was utilised instead of *B. subtilis* 168 wild-type to ensure recombination of the expression plasmids into the wild-type *aprE* locus did not occur and to ensure a single protein purification construct was obtained.

4.3.3.2 Validation of subtilisin E variant overexpression

To ensure pBAC0059/0060/0068/0069 yielded detectable levels of each secreted variant in the BAC0114 background, each strain was grown in the presence or absence of IPTG and the supernatant was analysed by SDS-PAGE (Figure 4.2). These were compared to a negative control strain containing the empty pHT01 plasmid. In the presence of IPTG a strong additional band is observed indicating the successful expression and secretion of each variant. Furthermore, BAC0114 was transformed with pBAC0057 and pBAC0058, expressing the native (*B. subtilis* 168) and triple variant (BAC0097) respectively with the native GTG start codon. Direct comparison of these strains to those with an ATG start codon shows the direct benefit of the altered start codon in obtaining higher levels of subtilisin E expression (Figure 4.2).

4.3.3.3 Size exclusion chromatography protein purification

Size exclusion chromatography (SEC) was utilised to purify the native subtilisin E, as well as each of the three variants. SEC separates proteins based on their size by isocratic elution, with larger proteins being eluted from the column before smaller proteins. Clarified and concentrated supernatant was loaded to the column and eluted (Section 2.12). The $UV_{280\ nm}$ absorbance of eluted proteins was monitored (Figure 4.3), and 5 mL fractions were collected throughout the purification process using the ÄKTA start protein purification system (GE healthcare). Following each purification, the fractions were analysed by SDS-PAGE to determine if the protein had been successfully isolated (Figure 4.3). In each case, fractions 16 and 17 yielded the pure subtilisin E variant, as observed in Figure 4.2, with no other proteins detectable by SDS-PAGE. As such the proteins were purified to a satisfactory level and could be characterised.

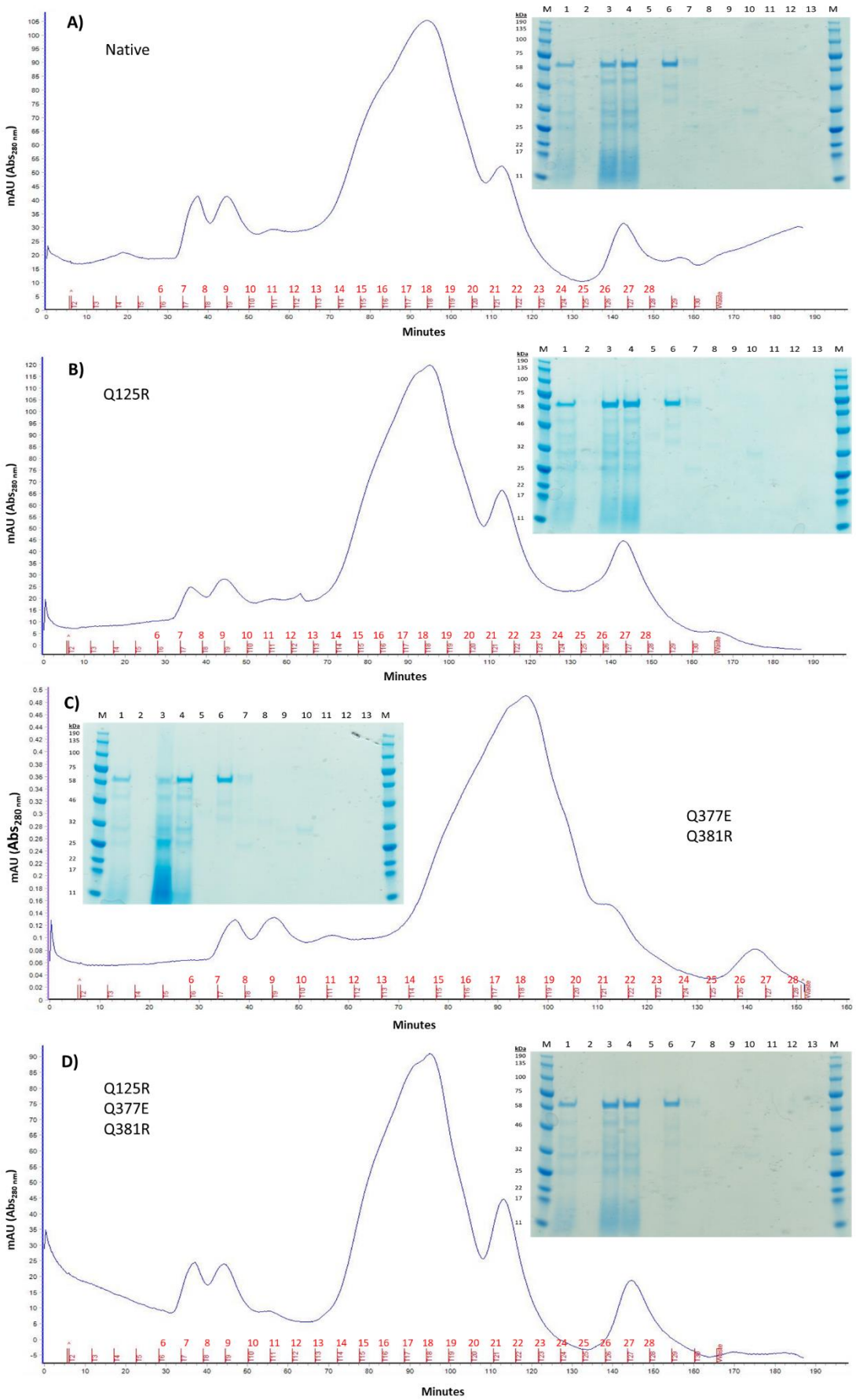

A)
Native
B)
Q125R
C)
Q377E
Q381R
D)
Q125R
Q377E
Q381R
mAU (Abs280 nm)
Minutes
M 1 2 3 4 5 6 7 8 9 10 11 12 13 M
kDa

Figure 4.3 – Subtilisin E variant purifications by SEC (previous page).

ÄKTA Purification chromatographs (Blue line = $UV_{280\ nm}$; Red lines and numbers = collection fractions) and their corresponding SDS-PAGE gels following purification of subtilisin E variants from strains A) BAC0119; B) BAC0121; C) BAC0122; D) BAC0120. Lanes on each SDS-PAGE gel represent: M – Blue Prestained Protein Standard, Broad Range (11-190 kDa) (New England Biolabs); 1 – clarified supernatant post dialysis; 2 – flow-through during concentration post dialysis; 3 – concentrated sample, pre filtration; 4 – concentrated sample, post filtration; 5 – fractions 6 & 7; 6 – fractions 8 & 9; 7 - fractions 10 & 11; 8 - fractions 12 & 13; 9 - fractions 14 & 15; 10 - fractions 16 & 17; 11 - fractions 18 & 19; 12 - fractions 20 & 21; 13 - fractions 22-28.

4.3.4 Thermotolerance of subtilisin E variants

The thermotolerance of the purified subtilisin E variants and native protein were analysed by the thermal shift assay (Figure 4.4)[132]. This assay utilises the SYPRO® Orange Protein Gel Stain (Merck) which non-specifically binds to hydrophobic residues. As the temperature increases in the reaction, the protein tertiary structure relaxes and exposes the hydrophobic core residues. This displaces the fluorescence quenching water molecules from the SYPRO® Orange molecule allowing fluorescence detection (excitation: 470 nm; emission: 570 nm) in a suitable instrument, such an RT-qPCR thermocycler[157]. Variants Q377E + Q381R and Q125R + Q377E + Q381R showed increased thermotolerance when compared to the native and Q125R proteins, confirming the importance of residues E377 and R381 for protein stability at higher temperatures. In contrast, the Q125R variant showed no increase in thermotolerance, most likely as there is no other mutated residue with which it can form a salt bridge.

The average T_m for each variant was established by performing a melt curve analysis of the thermal shift assay data (Figure 4.4). An increase in T_m of up to 1.4 °C was established and verified as statistically significant by use of an unpaired t test with Welch's correction (Figure 4.5). The greatest increase in T_m of 1.4 °C was found in the Q377E + Q381R variant (54.8 °C, P value = < 0.0001), while the full salt bridge triad variant showed an increase of 1.2 °C (54.6 °C, P value = < 0.0001). The Q125R variant showed no significant difference in T_m when compared to the WT (53.4 °C and 53.5 °C respectively).

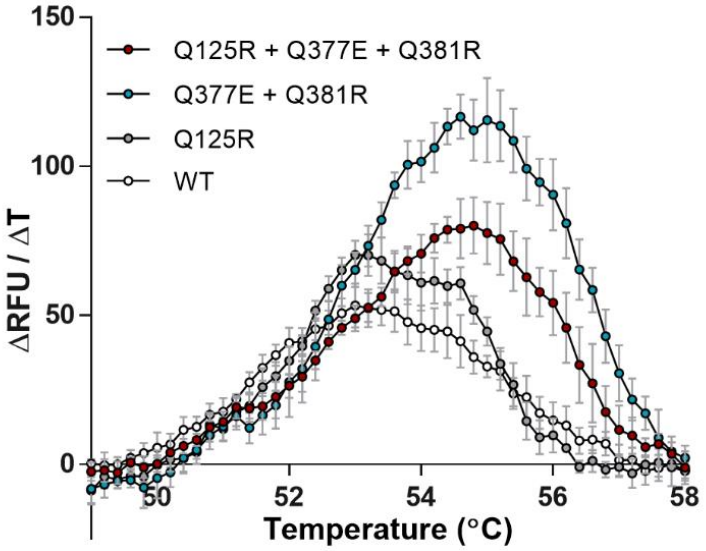


Figure 4.4 - Thermal shift assay of the thermostability of each subtilisin E variant and WT.

Increase in fluorescence was detected as hydrophobic regions of the protein were exposed as the protein denatured. The protein melting temperature of each variant and WT was calculated with a melt curve analysis, determining the peak rate of protein unfolding per temperature increase. Error bars indicate the standard deviation between six replicates.

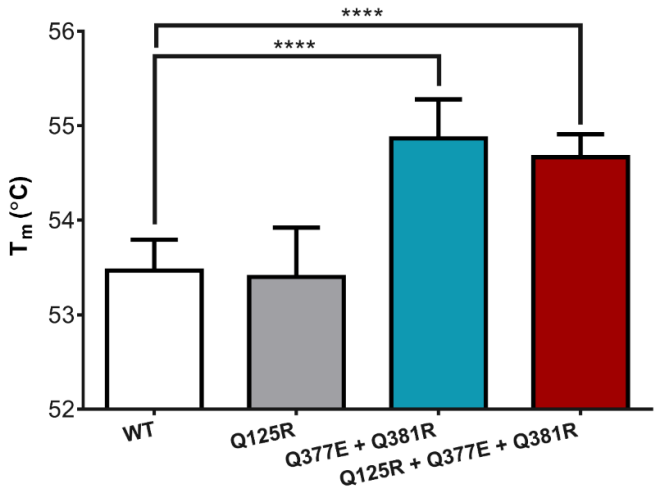


Figure 4.5 - Average T_m between Subtilisin variants and WT replicates.

The T_m of each subtilisin E variant and WT was calculated as the temperature at which the peak rate of protein unfolding was observed for each of six replicated. Error bars indicate the standard deviation. **** = p value summary (p = < 0.0001) following unpaired t test with Welch's correction.

4.3.5 Activity retention of subtilisin E variants

To ensure the introduced mutations had not negatively affected catalytic efficiency of the enzyme, the activity for each variant and the WT was established under neutral conditions (pH 7.5, 37 °C) by measuring the degradation of casein using Folin's reagent (Figure 4.6)[128]. Both the Q125R and Q125R + Q377E + Q381R variants were 60% less active when compared to the WT, suggesting that the Q125R residue is important for catalytic activity. The Q377E + Q381R variant on the other hand showed an increase in protease activity of 46.5% (P value = < 0.0001).

Residual enzyme activity was measured following incubation of the enzyme variants at 55 °C (Figure 4.7). Variant Q377E + Q381R was found to be the best at retaining its enzymatic activity when incubated for over 20 minutes. An improvement of 12.1% and 15.1% in activity relative to the WT following incubation for 40 and 60 minutes respectively was noted. The Q125R + Q377E + Q381R variant showed a 34.9% decrease in activity relative to the WT following 20 minutes incubation. Similarly, the Q125R variant showed a decrease in activity of 54.1% at 20 minutes incubation.

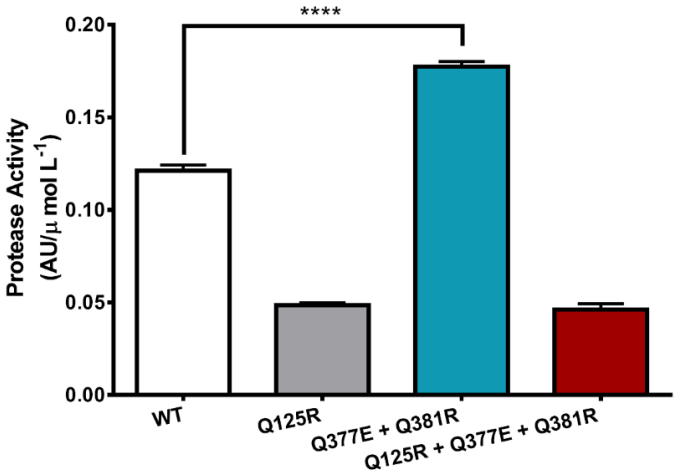


Figure 4.6 - Protease activity assay.

Protease activity at pH 7.5, 37 °C for 10 minutes. *N*ormalised for protein concentration determined by absorbance at 280 nm. Absorbance units (AU) determined at 660 nm. **** = p value summary (p = < 0.0001) following unpaired t test with Welch's correction. Error bars indicate standard deviation between triplicates.

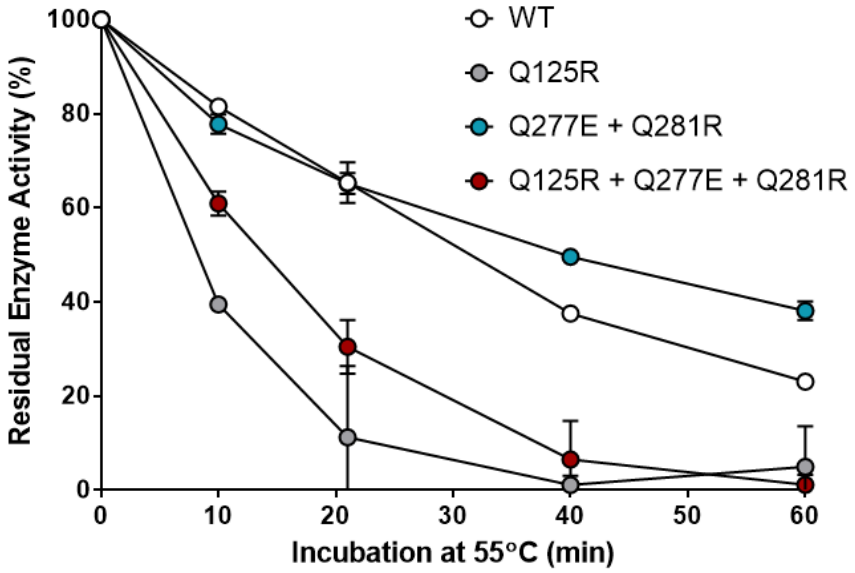


Figure 4.7 - Residual protease activity assay.

Residual protease activity under neutral conditions following incubation at 55 °C for various lengths of time, determined by a casein degradation enzymatic assay. Error bars indicate standard deviation between technical triplicates.

4.4 Discussion

In this chapter we have applied the CRISPR-Cas9 tools developed in Chapter 3 to perform rapid protein engineering. In under a week we carried out two rounds of *in situ* CRISPR-Cas9 editing of the subtilisin E gene, *aprE*. A salt bridge was introduced in two variants, Q377E + Q381R and Q125R + Q377E + Q381R. with the Q337E + Q381R variant showing an increase of 46.5% in subtilisin E activity, as well as a 1.4 °C increase in thermostability. To the best of our knowledge these modifications have not been combined before in *B. subtilis* subtilisin E.

Interestingly, the Q125R + Q377E + Q381R variant did not retain its activity levels to the same level as the Q377E + Q381R variant. Indeed, in both variants where the Q125R mutation is introduced a drastic drop in enzyme activity was observed (Figure 4.6 and Figure 4.7), indicating a vital catalytic activity role for residue Q125 within *B. subtilis* subtilisin E. A similar study of salt bridges in subtilisin E was performed by Erwin *et al*., where mutations Q125E + Q377E formed a salt

bridge based on X-ray crystal structure, but resulted in a 1.2 °C drop in thermotolerance[158]. Future analysis of the variants prepared in this study should also incorporate an empty vector negative control reaction which will ensure that the response observed in the protease activity assays were not due to a small contamination of a highly active protease.

This method for genomic modifications by CRISPR-Cas9 allows rapid, *in situ* protein engineering of industrially relevant strains. Furthermore, discovery and optimisation of molecular biology tools such as the work presented here, increases the speed and efficiency at which novel biocatalysts can be developed for sustainable bioprocesses. This method could be used for *in situ* directed evolution of protein regions through the construction of a dDNA library and the targeting of a single PAM site near the region of interest.

Chapter 5 – Transcriptional regulation via CRISPR tools in *Bacillus subtilis*

5.1 Introduction

The CRISPR-Cas9 toolbox for transcriptional interference (CRISPRi) within *B. subtilis* was first developed by Westbrook *et al*. through the use of chromosomally integrated *dcas9* and sgRNAs[106]. Latterly, Peters *et al*. validated their similarly integrated in the genome, dCas9 based CRISPRi system in *B. subtilis* as titratable for downregulation, sterically hindering the progression of the RNA polymerase (RNAP). This system was used to analyse all essential genes within *B. subtilis* 168[73]. Both these systems utilised chromosomal integrations of *dcas9* and sgRNA expression cassettes as it was thought that plasmid-based systems could be less stable. However, plasmids based on theta replication, such as pHT01, have been shown to be highly stable in *B. subtilis*[159]. Additionally, plasmid-based systems have a level of flexibility lacking in chromosomal integration methods. In cases where CRISPRi is utilised to identify targets for deletion, these integrated components would require removal prior to gene knock-out with the catalytically active Cas9 or other Cas9-based downstream applications. This is due to the risk of integration of Cas9 into dCas9 by homologous recombination.

Furthermore, multiplexing has been achieved through the use of individual sgRNA expression constructs[73]. Natural CRISPR-Cas9 systems utilise two RNA guiding components - tracrRNA and crRNA. Here we develop and exemplify a CRISPR system, compatible with *B. subtilis*, for the simultaneous multiplexed targeting of *amyE* and *gfpmut3* through the utilisation of a plasmid system where the sgRNA has been split to individually expressed tracrRNA, and the expression of a crRNA array to allow simple cloning of multiple targeting protospacers (Figure 5.1).

The transcriptional alteration systems described above only focus on the downregulation of gene expression by disruption of RNAP binding or RNA elongation. Transcriptional activation with CRISPR tools (CRISPRa) has been widely used in mammalian cell research, as well as in *S. cerevisiae*, and generally consist of a activation domain bound to dCas9 in order to attract the RNAP to the target gene[72]. No such dCas9 fusions had been reported within *B. subtilis*, in contrast to *E. coli*[59,160].

Bacterial RNAP (Figure 5.2) consists of 4 primary subunits: two α (RpoA) subunits; one β (RpoB) subunit; and one β' (RpoC) subunit. Combined, these form the $\alpha_2\beta\beta'$ complex which elongates RNA when it associates to a σ factor specific to the promoter elements upstream of a gene or operon[161,162]. Other small subunits, termed δ, ε, and ω are known to associate with the RNAP and aid in its assembly, increasing transcriptional specificity and influencing the recycling of RNAP[162].

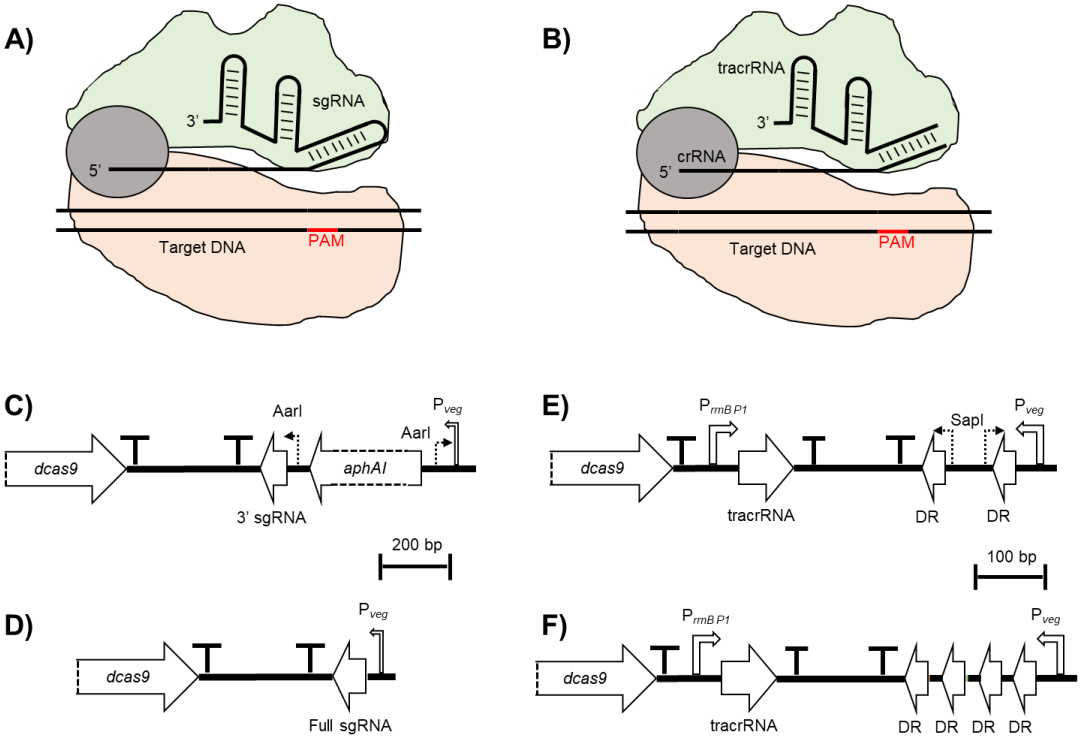


Figure 5.1 – Schematic diagram of CRISPR-Cas9 guiding RNA cloning sites.

Diagram showing the single guide RNA (sgRNA) (**A**) or tracrRNA-crRNA complex (**B**) bound to dCas9 enabling identification of the target PAM site. The elements of dCas9 are shown (recognition lobe – green; catalytically inactivated nuclease lobe – pink; catalytically inactivated HNH domain – grey). **C)** Diagram of the sgRNA cloning site. The kanamycin resistance gene (*aphAI* – not to scale) is removed by the type IIS restriction enzyme, AarI. A 20 bp protospacer region can then be introduced between the cleavage sites to complete the sgRNA (**D**). **E)** Diagram of the crRNA array cloning site. Protospacer regions, flanked by direct repeats (DR) are introduced (**F**) following cleavage with the type IIS restriction enzyme SapI. Dashed arrows indicate the orientation of DNA cleavage by AarI and SapI.

In *E. coli*, the ω subunit has been shown to associate with, and ensure correct folding of the β' subunit of RNAP (Figure 5.2). Furthermore, ω promotes the binding of β' with the $\alpha_2\beta$ complex[161,162]. *E. coli* and *Mycobacterium smegmatis* with *rpoZ* (ω) deleted were found to have misfolded and degraded β' subunits[161]. Dove and Hochschild found that the *E. coli* ω subunit could promote a 70-fold increase in transcription of a target genes, when fused to the DNA binding cI repressor from bacteriophage λ, within an *E. coli ΔrpoZ* strain[163]. Bikard *et al*. exemplified CRISPRa in *E. coli* through the use of dCas9 fused to the ω subunit (Figure 5.2)[59]. A maximum transcriptional increase of 23-fold was observed when ω was fused to the N-terminus of dCas9 (dCas9-ω) and targeted to the template DNA strand, 59 nt upstream of the -35 promoter element, controlling expression of a GFP reporter protein[59].

In this chapter we attempt to translate the CRISPRa method successfully exemplified in *E. coli* by Bikard *et al*. into *B. subtilis* 168[59], as well as attempting CRISPRa with an alternative transcriptional activator derived from the *B. subtilis* specific bacteriophage Φ29 protein 4, which is known to activate transcription of the viral genome at the A3 promoter through interactions with the RNAP α subunit (Figure 5.2).

Lastly, this chapter describes the first reported identification of a catalytically inactive variant of the MAD7 nuclease (isolated from *Eubacterium rectale* and also known as ErCas12a[71]), termed dMAD7. This tool is based on MAD7 which, according to Inscripta, is free to use for industrial and academic R&D[70]. MAD7 has been shown to be functional for genome editing within *E. coli*, *S. cerevisiae,* mouse and human HEK293T cells[67], zebrafish[71], as well as *B. subtilis* in this study (Section 3.3.2). MAD7, like other related Cpf1 (Cas12a) nucleases, has a T-rich PAM site preference of 5´-YTTN-3' (Y = T or C)[67] and natively utilises a single gRNA molecule, as opposed to the crRNA-tracrRNA complex utilised by Cas9 nucleases (Figure 5.1B).

We propose dMAD7 can be used to identify targets, which could subsequently be deleted from the chromosome using MAD7 enabled editing. The resulting strain, according to Inscripta[70], would not incur any read through royalties otherwise associated with Cas9 or Cpf1 nucleases, provided the final strain does not harbour the MAD7 or dMAD7 gene.

The work in this chapter relating to dMAD7 was published in February 2020 and a copy of this publication can be found in Appendix G.

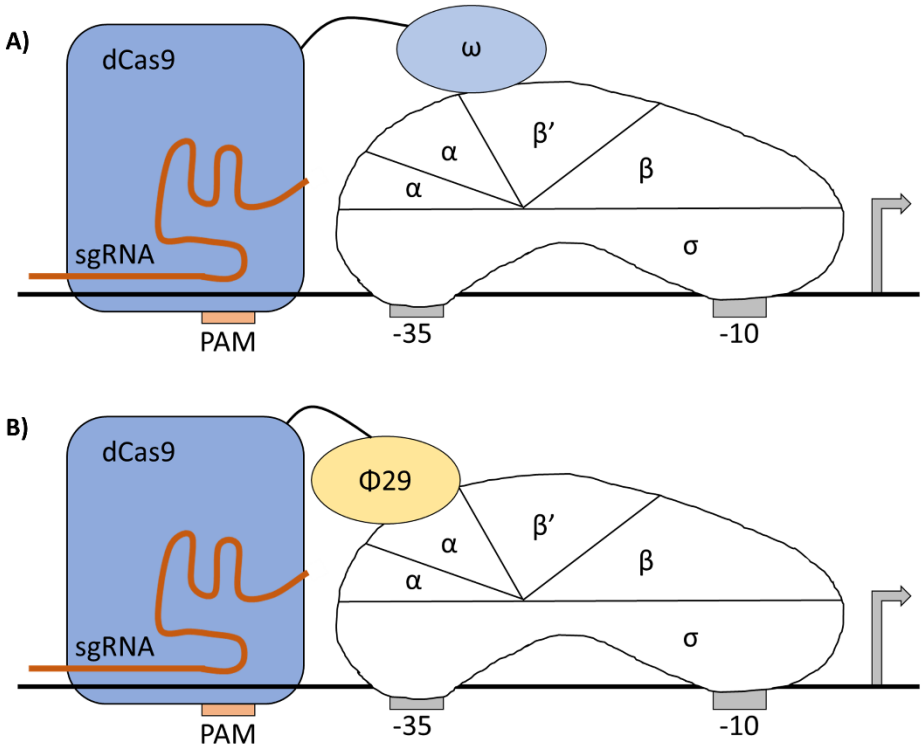


Figure 5.2 – Schematic representation of CRISPRa within bacteria.

A) the system employed by Bikard et al. where dCas9 is fused to the RNA polymerase (RNAP) ω subunit (known to interact with the β' subunit) to attract the RNAP complex (consisting of two α subunits, and one of each β, β' and σ subunits) to activate transcription at a specific promoter (-35 and -10) element. The subsequent transcriptional start site is indicated with an arrow. The complex is targeted to the correct site using a single guide RNA (sgRNA), containing a protospacer targeting DNA next to a protospacer adjacent motif (PAM) site. **B)** Similar to the system employed by Bikard et al., the bacteriophage Φ29 protein 4 is fused to dCas9 to interact with the RNAP α subunit and act as an alternative transcriptional activator.

5.2 Materials and methods

5.2.1 Plasmid and strain construction

The plasmids and strains used in this chapter are outlined below in Table 5.1. The oligonucleotides used in this chapter can be found in Appendix C.

Table 5.1 – Strains and plasmids used in this chapter.

Strain/Plasmid	Description/Genotype	Reference
Strains		
B. subtilis 168	*trpC2*	Laboratory stock
BAC0110	*ΔrpoZ*	This chapter.
BAC0111	*B. subtilis* 168 *ΔbglS::aph(3')-lIIa*-LIC site-*gfpmut3*	Section 3.2.1.2
BAC0205	*B. subtilis* 168 *ΔbglS::aph(3')-lIIa*-P_{liaG}-*gfpmut3*	This chapter.
BAC0206	BAC0110 *ΔbglS::aph(3')-lIIa*- P_{liaG}-*gfpmut3*	This chapter.
BAC0238	BAC0205 with pBAC0097	This chapter.
BAC0239	BAC0205 with pBAC0098	This chapter.
BAC0240	BAC0205 with pBAC0099	This chapter.
BAC0241	BAC0205 with pBAC0100	This chapter.
BAC0242	BAC0205 with pBAC0101	This chapter.
BAC0243	BAC0205 with pBAC0102	This chapter.
BAC0244	BAC0205 with pBAC0103	This chapter.
BAC0245	BAC0205 with pBAC0104	This chapter.
BAC0246	BAC0205 with pBAC0110	This chapter.
BAC0247	BAC0205 with pBAC0111	This chapter.

BAC0248	BAC0205 with pBAC0112	This chapter.
BAC0249	BAC0205 with pBAC0113	This chapter.
BAC0250	BAC0205 with pBAC0114	This chapter.
BAC0251	BAC0205 with pBAC0115	This chapter.
BAC0252	BAC0206 with pBAC0097	This chapter.
BAC0253	BAC0206 with pBAC0098	This chapter.
BAC0254	BAC0206 with pBAC0099	This chapter.
BAC0255	BAC0206 with pBAC0100	This chapter.
BAC0256	BAC0206 with pBAC0101	This chapter.
BAC0257	BAC0206 with pBAC0102	This chapter.
BAC0258	BAC0206 with pBAC0103	This chapter.
BAC0259	BAC0206 with pBAC0104	This chapter.
BAC0260	BAC0206 with pBAC0110	This chapter.
BAC0261	BAC0206 with pBAC0111	This chapter.
BAC0262	BAC0206 with pBAC0112	This chapter.
BAC0263	BAC0206 with pBAC0113	This chapter.
BAC0264	BAC0206 with pBAC0114	This chapter.
BAC0265	BAC0206 with pBAC0115	This chapter.
BAC0279	BAC0205 with pBAC0120	This chapter.
BAC0280	BAC0205 with pBAC0121	This chapter.
BAC0281	BAC0205 with pBAC0122	This chapter.
BAC0282	BAC0205 with pBAC0123	This chapter.

BAC0283	BAC0205 with pBAC0124	This chapter.
BAC0284	BAC0205 with pBAC0125	This chapter.
BAC0285	BAC0205 with pBAC0126	This chapter.
BAC0286	BAC0205 with pBAC0127	This chapter.
BAC0288	BAC0111 *ΔbglS::aph(3')-IIIa*-P_{veg}-*gfpmut3*	Section 3.2.1.2
BAC0289	BAC0111 *ΔbglS::aph(3')-IIIa*-$P_{rrnB\ P1}$-*gfpmut3*	This chapter.
BAC0295	BAC0288 with pBAC0132	This chapter.
BAC0296	BAC0288 with pBAC0133	This chapter.
BAC0297	BAC0288 with pBAC0104	This chapter.
BAC0298	BAC0288 with pBAC0102	This chapter.
BAC0299	BAC0288 with pBAC0103	This chapter.
BAC0300	BAC0288 with pBAC0092	This chapter.
BAC0301	BAC0288 with pBAC0093	This chapter.
BAC0302	BAC0288 with pBAC0094	This chapter.
BAC0303	BAC0288 with pBAC0095	This chapter.
BAC0304	BAC0288 with pBAC0096	This chapter.
BAC0305	BAC0289 with pBAC0133	This chapter.
BAC0306	BAC0289 with pBAC0104	This chapter.
BAC0307	BAC0289 with pBAC0102	This chapter.
BAC0308	BAC0289 with pBAC0103	This chapter.
BAC0309	BAC0289 with pBAC0092	This chapter.
BAC0310	BAC0289 with pBAC0093	This chapter.

BAC0311	BAC0289 with pBAC0094	This chapter.
BAC0312	BAC0289 with pBAC0095	This chapter.
BAC0313	BAC0289 with pBAC0096	This chapter.
BAC0314	BAC0289 with pBAC0134	This chapter.
BAC0315	BAC0289 with pBAC0135	This chapter.
BAC0330	BAC0288 with pBAC0164	This chapter.
BAC0331	BAC0288 with pBAC0168	This chapter.
BAC0348	BAC0288 with pBAC0194	This chapter.
BAC0349	BAC0288 with pBAC0195	This chapter.
BAC0350	BAC0288 with pBAC0189	This chapter.
BAC0351	BAC0288 with pBAC0190	This chapter.
BAC0352	*B. subtilis* 168 with pBAC0194	This chapter.
BAC0353	*B. subtilis* 168 with pBAC0184	This chapter.
BAC0354	*B. subtilis* 168 with pBAC0189	This chapter.
BAC0355	*B. subtilis* 168 with pBAC0188	This chapter.
BAC0360	*B. subtilis* 168 with pBAC0212	This chapter.
BAC0361	*B. subtilis* 168 with pBAC0213	This chapter.
BAC0362	*B. subtilis* 168 with pBAC0214	This chapter.
BAC0363	*B. subtilis* 168 with pBAC0215	This chapter.
BAC0364	BAC0288 with pBAC0207	This chapter.
BAC0365	BAC0288 with pBAC0208	This chapter.
BAC0366	BAC0288 with pBAC0209	This chapter.

BAC0367	BAC0288 with pBAC0210	This chapter.
BAC0368	BAC0288 with pBAC0211	This chapter.
BAC0377	BAC0288 with pBAC0222	This chapter.
BAC0378	BAC0288 with pBAC0220	This chapter.
BAC0380	BAC0288 with pBAC0219	This chapter.
BAC0381	BAC0288 with pBAC0212	This chapter.
BAC0382	BAC0288 with pBAC0215	This chapter.
Plasmids		
pdCas9-bacteria	*tetR*; *dCas9* (*S. pyogenes*); rrnB T1; p15a ori; *cat*	[138]
pGFPbglS	*bla*; *5´ bglS*; *aph(3')-IIIa*; LIC site; *gfpmut3*; *3' bglS*	[124]
pMK-RQ-dMAD7	ColE1 ori; *aphAI*; *dMAD7* codon optimised for *B. subtilis*	Synthesised by Thermo Fisher Scientific
pMK-RQ-tracrRNA-crRNA array	ColE1 ori; *aphAI*; pHT01 MCS; *trpA* terminator; $P_{rrnB\ P1}$; tracrRNA; T7 terminator; *trpA* terminator; DR; SapI sites; DR; P_{veg}; PmlI site	Synthesised by Thermo Fisher Scientific
pWJ66	*cat*; p15A ori; tracrRNA; *dcas9-ω(Ec)*; crRNA cloning site	[59]
pBAC0001	pHT01 with SapI sites removed	Section 3.2.1.1
pBAC0008	pBAC0001 with rrnB T1 T2; sgRNA Cas9 handle; protospacer cloning site; P_{veg}	Section 3.2.1.1
pBAC0015	pBAC0008 with *cas9* (*S. pyogenes*). Cas9 expression regulated by the P_{grac}	Section 3.2.1.1

pBAC0018	pBAC0008 with *dcas9* (*S. pyogenes*). *dcas9* expression regulated by the P_{grac}	This chapter.
pBAC0035	pBAC0015 with non-targeting sgRNA DNA	Section 3.2.1.1
pBAC0041	pBAC0015 with sgRNA targeting 25 bp downstream of the start codon within *B. subtilis* 168 *amyE*	Section 3.2.1.1
pBAC0065	pBAC0015 with sgRNA targeting 143 bp downstream of the start codon within *B. subtilis* 168 *rpoZ*	This chapter.
pBAC0067	pBAC0008 with dcas9-ω(Ec). dcas9-ω(Ec) expression regulated by the P_{grac}	This chapter.
pBAC0082	pBAC0018 with dcas9-ω(Bs). dcas9-ω(Bs) expression regulated by the P_{grac}	This chapter.
pBAC0090	pBAC0018 with dcas9-Φ29. dcas9-Φ29 expression regulated by the P_{grac}	This chapter.
pBAC0091	pGFPbglS with P_{liaG} inserted at LIC site	This chapter.
pBAC0092	pBAC0082 with sgRNA targeting 113 bp upstream of the P_{veg} and $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0288/0289	This chapter.
pBAC0093	pBAC0082 with sgRNA targeting 131 bp upstream of the P_{veg} and $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0288/0289	This chapter.
pBAC0094	pBAC0082 with sgRNA targeting 139 bp upstream of the P_{veg} and $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0288/0289	This chapter.
pBAC0095	pBAC0082 with sgRNA targeting 163 bp upstream of the P_{veg} and $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0288/0289	This chapter.

pBAC0096	pBAC0082 with sgRNA targeting 200 bp upstream of the P_{veg} and $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0288/0289	This chapter.
pBAC0097	pBAC0067 with sgRNA targeting 105 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0098	pBAC0067 with sgRNA targeting 125 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0099	pBAC0067 with sgRNA targeting 145 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0100	pBAC0067 with sgRNA targeting 154 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0101	pBAC0067 with non-targeting sgRNA DNA	This chapter.
pBAC0102	pBAC0082 with sgRNA targeting 80 bp upstream of the P_{liaG}, and 16 bp upstream P_{veg} and $P_{rrnB\ P1}$, -35 promoter elements for *gfpmut3* expression in BAC0205/0206/0288/0289	This chapter.
pBAC0103	pBAC0082 with sgRNA targeting 80 bp upstream of the P_{liaG}, and 16 bp upstream P_{veg} and $P_{rrnB\ P1}$, -35 promoter elements for *gfpmut3* expression in BAC0205/0206/0288/0289	This chapter.
pBAC0104	pBAC0082 with non-targeting sgRNA DNA	This chapter.
pBAC0110	pBAC0090 with sgRNA targeting 80 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.

pBAC0111	pBAC0090 with sgRNA targeting 97 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0112	pBAC0090 with sgRNA targeting 125 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0113	pBAC0090 with sgRNA targeting 145 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0114	pBAC0090 with sgRNA targeting 154 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205/0206	This chapter.
pBAC0115	pBAC0090 with non-targeting sgRNA DNA	This chapter.
pBAC0120	pBAC0082 with sgRNA targeting 84 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.
pBAC0121	pBAC0082 with sgRNA targeting 73 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.
pBAC0122	pBAC0082 with sgRNA targeting 68 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.
pBAC0123	pBAC0082 with sgRNA targeting 56 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.
pBAC0124	pBAC0082 with sgRNA targeting 45 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.

pBAC0125	pBAC0082 with sgRNA targeting 30 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.
pBAC0126	pBAC0082 with sgRNA targeting 84 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.
pBAC0127	pBAC0082 with sgRNA targeting 11 bp upstream of the P_{liaG} -35 promoter element for *gfpmut3* expression in BAC0205	This chapter.
pBAC0129	pBAC0015 with sgRNA targeting 18 bp upstream of the start codon within BAC0111 *gfpmut3*.	This chapter.
pBAC0132	pBAC0082 with sgRNA targeting 16 bp upstream of the P_{veg} -35 promoter element for *gfpmut3* expression in BAC0288	This chapter.
pBAC0133	pBAC0082 with sgRNA targeting 61 bp upstream of the P_{veg} and $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0288/0289	This chapter.
pBAC0134	pBAC0082 with sgRNA targeting 20 bp downstream of the $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0289	This chapter.
pBAC0135	pBAC0082 with sgRNA targeting 9 bp upstream of the $P_{rrnB\ P1}$ -35 promoter element for *gfpmut3* expression in BAC0289	This chapter.
pBAC0155	pBAC0001 with rrnB T1 T2; T7 terminator; protospacer cloning site; gRNA MAD7 handle; P_{veg}	Section 3.2.1.1
pBAC0157	pBAC0001 with tracrRNA-crRNA array cloned between BamHI and SmaI sites	This chapter.
pBAC0158	pBAC0155 with *MAD7* cloned between BamHI and XbaI sites.	Section 3.2.1.1

pBAC0162	pBAC0158 with gRNA targeting 21 bp downstream of the start codon within *B. subtilis* 168 *amyE*	Section 3.2.1.1
pBAC0163	pBAC0158 with non-targeting gRNA DNA	Section 3.2.1.1
pBAC0164	pBAC0157 with *dcas9* cloned between BamHI and XbaI sites	This chapter.
pBAC0165	pBAC0015 with sgRNA targeting 27 bp downstream of the start codon within BAC0288 *gfpmut3*	Section 3.2.1.1
pBAC0166	pBAC0158 with sgRNA targeting 21 bp downstream of the start codon within BAC0288 *gfpmut3*	Section 3.2.1.1
pBAC0168	pBAC0164 with crRNA array DNA targeting (in order) 47 bp downstream of the start codon within BAC0288 *amyE* and 68 bp upstream of the start codon within BAC0288 *gfpmut3*	This chapter.
pBAC0184	pBAC0041 with *dcas9* replacing *cas9*	This chapter.
pBAC0187	pBAC0155 with *dMAD7* from pMK-RQ-dMAD7. *dMAD7* expression regulated by P_{grac}	This chapter.
pBAC0188	pBAC0162 with *dMAD7* replacing *MAD7*	This chapter.
pBAC0189	pBAC0163 with *dMAD7* replacing *MAD7*	This chapter.
pBAC0190	pBAC0166 with *dMAD7* replacing *MAD7*	This chapter.
pBAC0194	pBAC0035 with *dcas9* replacing *cas9*	This chapter.
pBAC0195	pBAC0165 with *dcas9* replacing *cas9*	This chapter.
pBAC0207	pBAC0187 with sgRNA targeting 9 bp downstream of the start codon within BAC0288 *gfpmut3*	This chapter.
pBAC0208	pBAC0187 with sgRNA targeting 16 bp downstream of the start codon within BAC0288 *gfpmut3*	This chapter.

pBAC0209	pBAC0187 with sgRNA targeting 43 bp downstream of the start codon within BAC0288 *gfpmut3*	This chapter.
pBAC0210	pBAC0187 with sgRNA targeting 78 bp downstream of the start codon within BAC0288 *gfpmut3*	This chapter.
pBAC0211	pBAC0187 with sgRNA targeting 80 bp downstream of the start codon within BAC0288 *gfpmut3*	This chapter.
pBAC0212	pBAC0187 with sgRNA targeting 4 bp downstream of the start codon within *B. subtilis* 168 *amyE*	This chapter.
pBAC0213	pBAC0187 with sgRNA targeting 11 bp downstream of the start codon within *B. subtilis* 168 *amyE*	This chapter.
pBAC0214	pBAC0187 with sgRNA targeting 27 bp downstream of the start codon within *B. subtilis* 168 *amyE*	This chapter.
pBAC0215	pBAC0187 with sgRNA targeting 51 bp downstream of the start codon within *B. subtilis* 168 *amyE*	This chapter.
pBAC0219	pBAC0187 with crRNA array DNA targeting (in order) 80 bp and 21 bp downstream of the start codon within BAC0288 *gfpmut3*	This chapter.
pBAC0220	pBAC0187 with crRNA array DNA targeting (in order) 80 bp downstream of the start codon within BAC0288 *gfpmut3* and 4 bp downstream of the start codon within BAC0288 *amyE*	This chapter.
pBAC0222	pBAC0187 with crRNA array DNA targeting (in order) 4 bp and 51 bp downstream of the start codon within BAC0288 *amyE*	This chapter.

LIC = Ligation Independent Cloning; MCS = Multiple Cloning Site; DR = direct repeat.

5.2.1.1 Plasmid construction

The construction of plasmids pBAC0008/0155 using the inABLE® assembly technique is described in detailed in Section 3.2.1.1. The introduction of *cas9* into

pBAC0008 (yielding pBAC0015) and *MAD7* into pBAC0155 (yielding pBAC0158) is also described in detailed in Section 3.2.1.1.

The catalytically inactive *dcas9* gene was PCR amplified (Section 2.5.5.1) from pdCas9-bacteria[138] with primers oMAP0073/0074, introducing a BsaI site and XbaI restriction enzyme recognition sites at the 5´ and 3' end of the gene, respectively. The amplified *dcas9* was digested (Section 2.5.8) with BsaI and XbaI, and ligated (Section 2.5.10) with the BamHI-HF and XbaI digested pBAC0008 backbone, yielding pBAC0018. BsaI was used due to *S. pyogenes dcas9* containing a BamHI recognition site.

The *dcas9-ω(Ec)* fusion was similarly cloned into the BamHI and XbaI digested pBAC0008 backbone following PCR amplification from pWJ66, yielding plasmid pBAC0067[59]. pWJ66 was a gift from Luciano Marraffini (Addgene plasmid #46570).

pBAC0082 carrying the *dcas9-ω(Bs)* fusion was cloned by PCR amplification of the 3' region of dCas9 from pBAC0018 with oMAP0223 (introducing a PmlI restriction enzyme recognition site) and oMAP0214. In parallel, the *B. subtilis rpoZ* gene (encoding the *B. subtilis* RNA polymerase ω subunit) was PCR amplified from *B. subtilis* 168 gDNA (gDNA preparation described in section 2.5.3) with primers oMAP0215 (introducing a 20 bp region of homology with oMAP0214) and oMAP0216 (introducing an XbaI restriction enzyme recognition site). These two PCR products were fused using OE-PCR (Section 2.5.5.3) with primers oMAP0216/0223, and the subsequent product was digested with PmlI and XbaI. pBAC0018 was also digested with PmlI and XbaI and ligated with the digested 3' *dCas9-rpoZ* OE-PCR product.

pBAC0090 carrying the *dCas9-Φ29* fusion protein was cloned in the same manner as pBAC0082 with the 3' region of dCas9 PCR amplified with the same primers. The genome sequence of Φ29 is available (NC_011048.1) and as such P4 was synthesised (codon optimised for *B. subtilis*) as linear dsDNA by Twist Bioscience (USA). *Φ29P4* was PCR amplified using primers oMAP0317/0318.

pBAC0187 was constructed by digesting pBAC0155 and pMK-RQ-dMAD7 (*dMAD7* (*MAD7* gene with D877A (codon `GCT` to `GAT`), E962A (codon `GCA` to `GAA`), D1213A (codon `GCT` to `GAT`) modifications) flanked by BamHI and XbaI recognition sites, synthesised by Thermo Fisher Scientific), with BamHI-HF and XbaI. The *dMAD7* gene fragment was subsequently ligated with the pBAC0155 backbone.

Plasmids used for Cas9-mediated editing were prepared from the pBAC0015 parental plasmid using phosphorylated and annealed oligonucleotide pairs as described in section 2.5.4.1. Plasmids used for analysis of the effector fusion proteins dCas9-ω(Ec), dCas9-ω(Bs) or dCas9-Φ29 were prepared from the pBAC0067, pBAC0082 or pBAC0090 parental plasmids respectively, using phosphorylated and annealed oligonucleotide pairs as described in section 2.5.4.1. Plasmids used for dMAD7 CRISPRi analysis were prepared from the pBAC0158 parental plasmid using phosphorylated and annealed oligonucleotide pairs as described in section 2.5.4.1. Information on the oligonucleotides used in construction, target gene and PAM site for the following CRISPR plasmids used in this chapter can be found in Table 2.12: pBAC0035; pBAC0041; pBAC0065; pBAC0097; pBAC0098; pBAC0099; pBAC0100; pBAC0101; pBAC0092; pBAC0093; pBAC0094; pBAC0095; pBAC0096; pBAC0102; pBAC0103; pBAC0104; pBAC0120; pBAC0121; pBAC0122; pBAC0123; pBAC0124; pBAC0125; pBAC0126; pBAC0127; pBAC0129; pBAC0132; pBAC0133; pBAC0134; pBAC0135; pBAC0110; pBAC0111; pBAC0112; pBAC0113; pBAC0114; pBAC0115; pBAC0162; pBAC0163; pBAC0165; pBAC0166; pBAC0207; pBAC0208; pBAC0209; pBAC0210; pBAC0211; pBAC0212; pBAC0213; pBAC0214; pBAC0215; pBAC0219; pBAC0220; pBAC0222.

pBAC0157 was constructed by ligating the BamHI and SmaI cut pBAC0001 with the BamHI and PmlI cut pMK-RQ-tracrRNA-crRNA array (Thermo Fisher Scientific). pBAC0157 contains a tracrRNA expression cassette, controlled by $P_{rrnB\ P1}$, and a crRNA array expression cassette, controlled by P_{veg}, and retained the P_{grac} controlled expression site successfully used for *cas9* expression (Section 3.3.1.2). *dcas9* was inserted following PCR amplification using oMAP0073/0074, digesting this with BsaI and XbaI, and ligating with pBAC0157 cut with BamHI and XbaI, yielding pBAC0164. pBAC0168 was designed to simultaneously target *gfpmut3* and *amyE*. A third, non-targeting crRNA was also included to ensure maturation of the first two crRNAs. The crRNA array to yield these targeting molecules was split between two sets of oligonucleotides due to their length. Each set, once phosphorylated and annealed (Section 2.5.4.1), contained a 3 nt overhang for ligation with the SapI digested pBAC0164 backbone, as well as a 4 nt overhang for ligation to the other oligonucleotide set. The three fragments were combined in a single ligation reaction. SapI replaced AarI in plasmids derived from pBAC0008 due to it being a more reliable, stable restriction enzyme with type IIS properties. SapI

could be used here since the crRNA array cloning region was synthesised, and the InABLE® DNA assembly method was not utilised (in which SapI digestion is generally required, preventing the downstream use of SapI).

pBAC0184/0194/0195 were prepared by digesting pBAC0041/0035/0165 respectively with BamHI-HF and XbaI to remove *cas9* and subsequently ligating these backbones with *dcas9* which was PCR amplified from pdCas9-bacteria with primers oMAP0073/0074. This PCR product was digested with BsaI and XbaI and ligated with the prepared backbones.

pBAC0188/0189/0190 were prepared by digesting pBAC0162/0163/0166 respectively with BamHI-HF and XbaI to remove *MAD7* and subsequently ligating these backbones with *dMAD7* which was isolated from pMK-RQ-dMAD7 with the same restriction enzymes.

Ligation independent cloning (LIC) was used for the introduction of P_{liaG} for *gfpmut3* expression on the plasmid pGFPbglS[124]. The method outlined by Bisicchia *et al*. was followed[124]. Oligonucleotides oMAP0244/0245 were used to PCR amplify P_{liaG} from *B. subtilis* 168 gDNA and yielded the final integration plasmid pBAC0091.

5.2.1.2 Strain construction

Using the approach described in section 2.7, CRISPR-Cas9 mediated chromosome editing was used to delete *rpoZ* from *B. subtilis* 168. dDNA OE-PCR products, amplified from *B. subtilis* 168 gDNA with oligonucleotide set oMAP0210/0211/0212/0213 was used alongside pBAC0065. Cas9/sgRNA plasmid was removed from the edited strain as described in section 2.7.6, yielding BAC0110. The deletion of *rpoZ* was verified by PCR amplification from gDNA with oMAP0210/0213 and D5000 Tapestation analysis as described in section 2.5.6.2.

Similarly, CRISPR-Cas9 mediated chromosome editing was used to insert the P_{veg} or $P_{rrnB\ P1}$ promoters upstream of the *gfpmut3* gene in BAC0111 (Section 3.2.1.2) to construct stain BAC0288 (Section 3.2.1.2) and BAC0289 respectively. To construct BAC0289, dDNA OE-PCR products, amplified from BAC0111 gDNA with oligonucleotide set oMAP0388/0393/0396/0397 was used alongside pBAC0129. The CRISPR-Cas9/sgRNA plasmid was removed from the edited strain as described in section 2.7.6. The insertion of $P_{rrnB\ P1}$ was verified by PCR amplification from BAC0289 gDNA (Section 2.5.3) with oMAP0803/0804 and sequencing.

Furthermore, GFPmut3 was detected by fluorescence emission analysis using Safe Imager 2.0 Blue Light Transilluminator and Amber Filter (Thermo Fisher Scientific) system (excitation: 470 nm, emission: 530 nm).

All remaining strains described in Table 5.1 were prepared by making their parental strain naturally competent (section 2.4.2) and transforming with 200 ng of the plasmid also described in Table 5.1.

5.3 Results

5.3.1 Evaluation of CRISPRa within *B. subtilis*

5.3.1.1 Effector protein selection

Following the successful exemplification of CRISPRa in *E. coli ΔrpoZ* using a dCas9 fusion to the ω subunit of this organism (ω(Ec)), this dCas9-effector combination was chosen to attempt CRISPRa within *B. subtilis* alongside two other effector proteins. These were the ω subunit (RpoZ) of RNAP in *B. subtilis* (ω(Bs)), and the Φ29 bacteriophage protein P4, known to interact with the α subunit of RNAP of *B. subtilis* to activate transcription of the viral genome at the viral A3 promoter[59,164]. It was hypothesised that this interaction could be harnessed for transcriptional activation through interactions between the Φ29 subunit and the native RNAP α subunit.

Since Bikard *et al*. had found the C-terminal fusion variant had yielded the highest CRISPRa effect (2.8-fold increase (dCas9-ω) compared to ~1.8-fold increase (ω-dCas9)), our design for dCas9-ω(Bs) and dCas9-Φ29 followed the same approach[59]. Similarly, the Ala-Ala linker region between dCas9 and ω(Ec), used by Bikard *et al*., was retained in our new dCas9-effector combinations.

5.3.1.2 Fluorescence-based transcriptional regulation screen

Bikard *et al*. observed CRISPRa to be functional in *E. coli* only when *rpoZ* had been deleted[59]. The WT strain had not been tested in parallel, but the assumption was carried through from results by Dove and Hochschild where transcriptional activation in *E. coli,* with an alternative DNA binding protein fused to RpoZ, was only observed when *rpoZ* was deleted[163]. We utilised our CRISPR-Cas9

genome editing method to delete *rpoZ* from the chromosome of *B. subtilis* 168, yielding BAC0110.

GFPmut3, a variant of the jellyfish *Aequorea victoria* GFP (containing mutations S65G and S72A) with an increased fluorescence intensity and a 106 nm shifted excitation maxima[165], was selected as the reporter for screening of CRISPRa candidates due to its ease of detection by fluorescence following excitation by blue light (470 nm) and its prior use within *B. subtilis*[113,124]. P_{liaG} is known to be a constitutive promoter[139]. Ligation independent cloning (LIC – section 2.5.13) was utilised for the insertion of P_{liaG} upstream of *gfpmut3* in pGFPbglS[124]. The resulting plasmid, pBAC0091, was recombined onto the chromosome at the *bglS* locus of both the WT *B. subtilis* 168 and BAC0110 strains, yielding strains BAC0205 and BAC0206 respectively following validation of the site-specific integration by colony PCR. Furthermore, integrants were identified by growth on an LB agar supplemented with kanamycin as the resistance gene (*aph(3')-IIIa*) was also chromosomally integrated upstream of *gfpmut3*. A further two background strains were prepared where *gfpmut3* expression was under control of the stronger promoters P_{veg} and $P_{rrnB\ P1}$, strains BAC0288 and BAC0289 respectively. The expression levels of *gfpmut3* under the control of $P_{liaG,}$ P_{veg} and $P_{rrnB\ P1}$ are shown in Figure 5.3. P_{liaG} was taken forward for initial CRISPRa analysis as it would allow the easiest identification of when CRISPRa is most effective since a further increase in expression by stronger promoters may not be clearly detectable.

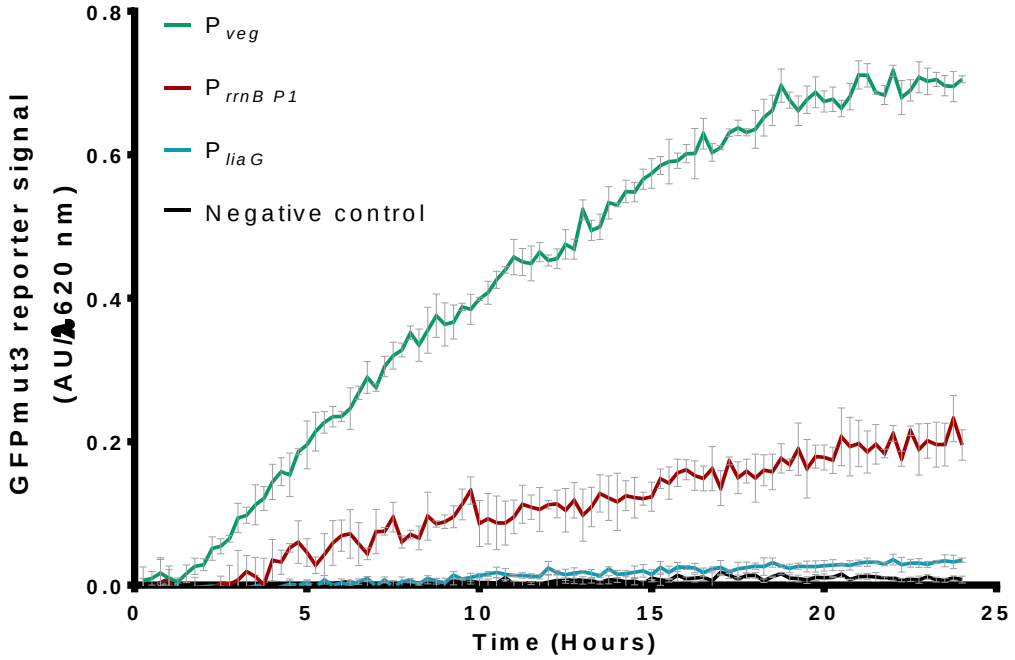


Figure 5.3 – Fluorescence signal of GFPmut3 transcriptional reporters.

Levels of fluorescence signal with strong (P_{veg} – BAC0288), medium ($P_{rrnB\ P1}$ – BAC0289) and weak (P_{liaG} – BAC0205) promoters controlling the expression of *gfpmut3*. The negative control strain was unaltered *B. subtilis* 168. Error bars indicate the standard deviation between three biological replicates.

To screen the selected effector proteins for CRISPRa activity by GFPmut3 transcriptional increase, various sgRNAs were designed varying in the distance between the PAM site and the -35 element of the promoter: 80 bp; 97 bp; 105 bp; 129 bp; 145 bp; 154 bp.

Six PAM sites were selected between 80 and 154 bp upstream of the P_{liaG} -35 element for targeting with the dCas9-effector protein fusions. In addition, a non-targeting plasmid was prepared for each dCas9-effector protein fusion to act as a negative control for CRISPRa.

Of the 21 designed dCas9-effector/sgRNA plasmids, 14 were successfully constructed. These CRISPRa targeting and non-targeting plasmids were used to transform BAC0205 (WT with P_{liaG}-*gfpmut3*) and BAC0206 (*ΔrpoZ* with P_{liaG}-*gfpmut3*) background strains. These 28 strains (BAC0238-BAC0265) were subsequently analysed for GFPmut3 fluorescence (section 2.10.2) in order to determine if the targeted dCas9-effector fusions caused an increase in GFPmut3

fluorescence levels (Figure 5.4, Figure 5.5 and Figure 5.6). Also included in each plate, was a positive control strain where *gfpmut3* was introduced onto the pHT01 plasmid and expressed by IPTG induced P_{grac} to ensure GFPmut3 could be detected (data not shown).

The results indicate the dCas9-ω(Bs) fusion protein yielded an increase of 91.3% in GFPmut3 fluorescence, relative to the non-targeting control following 20 hours of growth, when targeting the PAM site 80 nt upstream of the -35 element in strain BAC0243 (Figure 5.5). The effect can be seen in the mean of the replicates from 8 hours into the analysis, with the effect becoming more pronounced following 15 hours of growth. The same increase in fluorescence was not observed in the *ΔrpoZ* background strain. No detectable CRISPRa effect was observed for the dCas9-ω(Ec) and dCas9-Φ29 fusion proteins under the tested conditions, relative to the non-targeting control strains (Figure 5.4 and Figure 5.6).

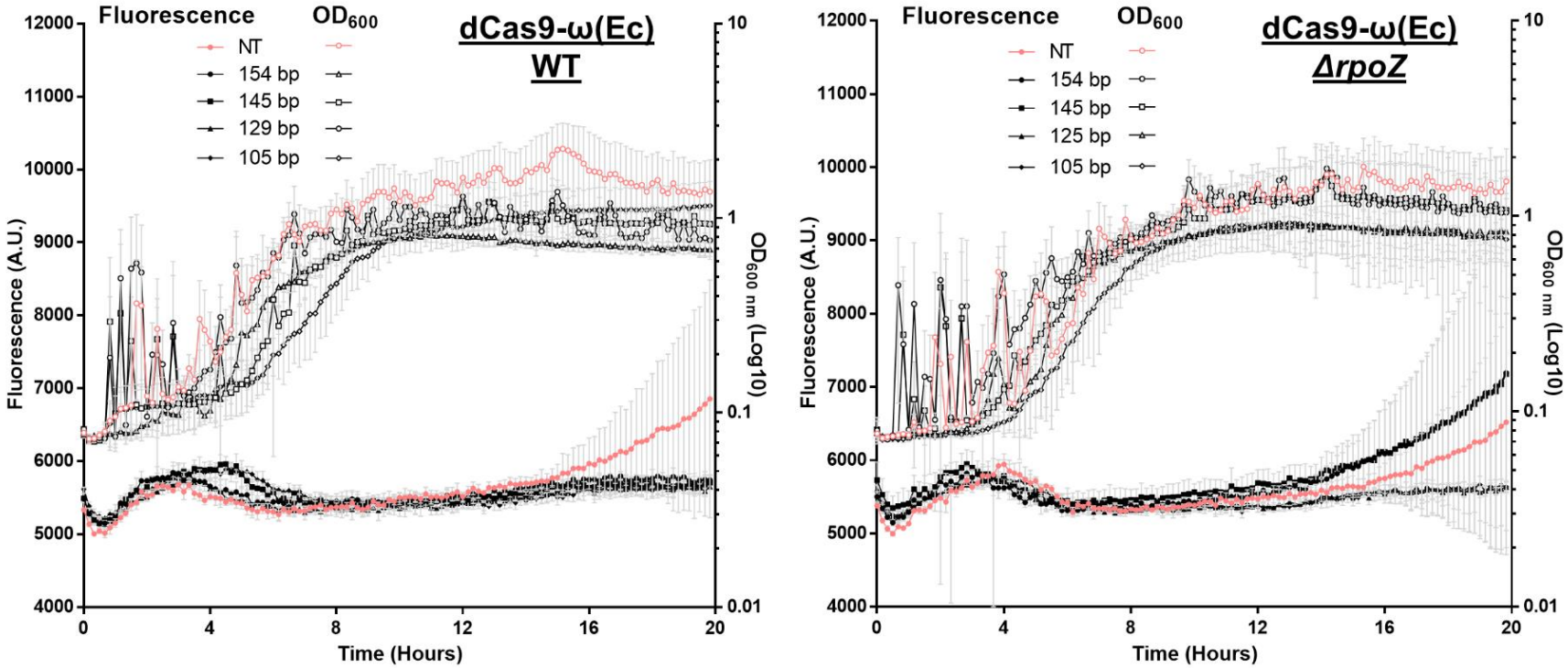


Figure 5.4 – Screening of a dCas9-ω(Ec) enabled CRISPRa system based on GFPmut3 transcriptional reporters.

Levels of fluorescence (solid symbols) and OD $_{600\ nm}$ (hollow symbols). *gfpmut3* expression is controlled by P_{liaG}. In either the 'WT' BAC0205 (left) or a *ΔrpoZ* BAC0206 (right) background strains, a plasmid is present expressing the candidate CRISPRa dCas9-effector fusion protein, dCas9-ω(Ec). Each plasmid expressed an sgRNA targeting PAM sites at the indicated distance from the 5´ of the -35 element of P_{liaG}. A non-targeting (NT) sgRNA control strain for each background strain was included (pink). All screened sgRNA targeted PAM sites on the sense strand. Error bars indicate the standard deviation between three biological replicates.

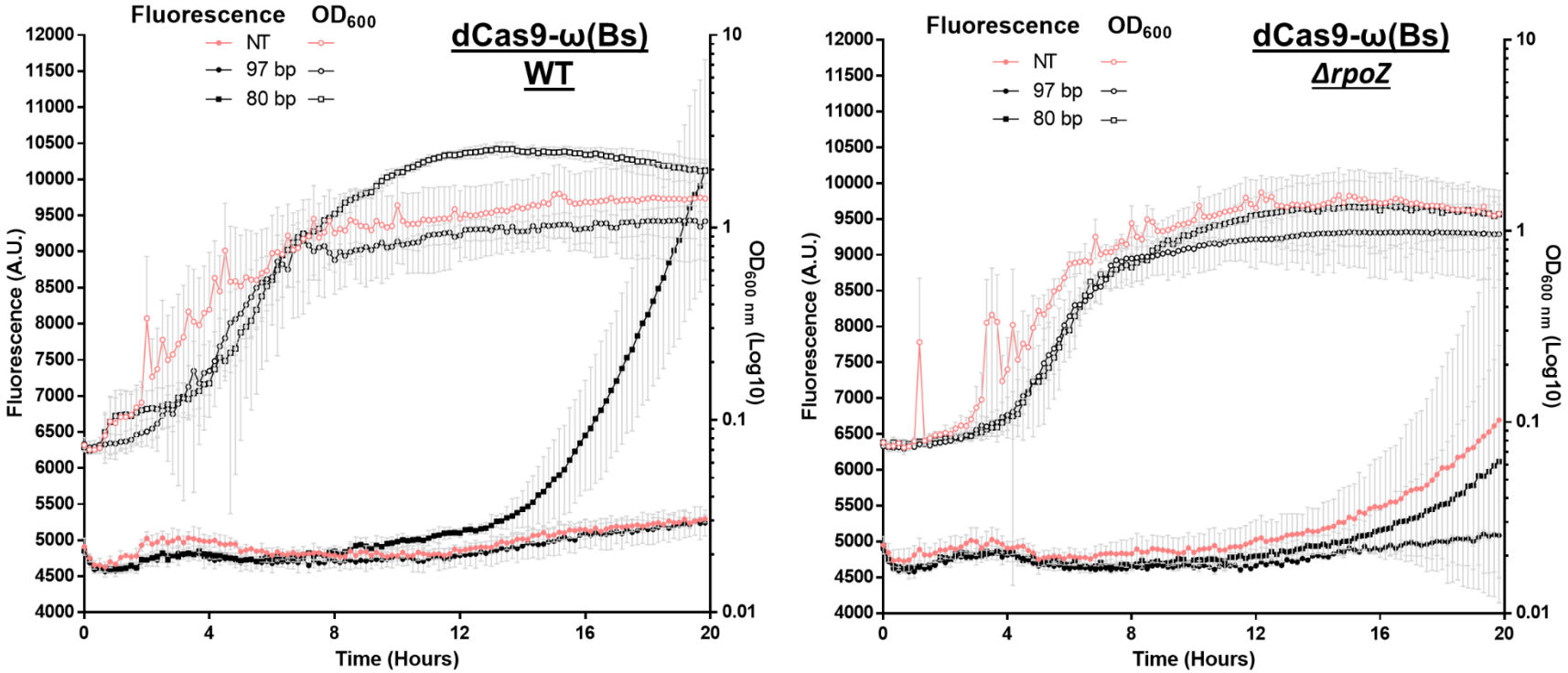


Figure 5.5 – Screening of a dCas9-ω(Bs) enabled CRISPRa system based on GFPmut3 transcriptional reporters.

Levels of fluorescence (solid symbols) and OD $_{600\ nm}$ (hollow symbols). *gfpmut3* expression is controlled by P_{liaG}. In either the 'WT' BAC0205 (left) or a *ΔrpoZ* BAC0206 (right) background strains, a plasmid is present expressing the candidate CRISPRa dCas9-effector fusion protein, dCas9-ω(Bs). Each plasmid expressed an sgRNA targeting PAM sites at the indicated distance from the 5´ of the -35 element of P_{liaG}. A non-targeting (NT) sgRNA control strain for each background strain was included (pink). All screened sgRNA targeted PAM sites on the sense strand. Error bars indicate the standard deviation between three biological replicates.

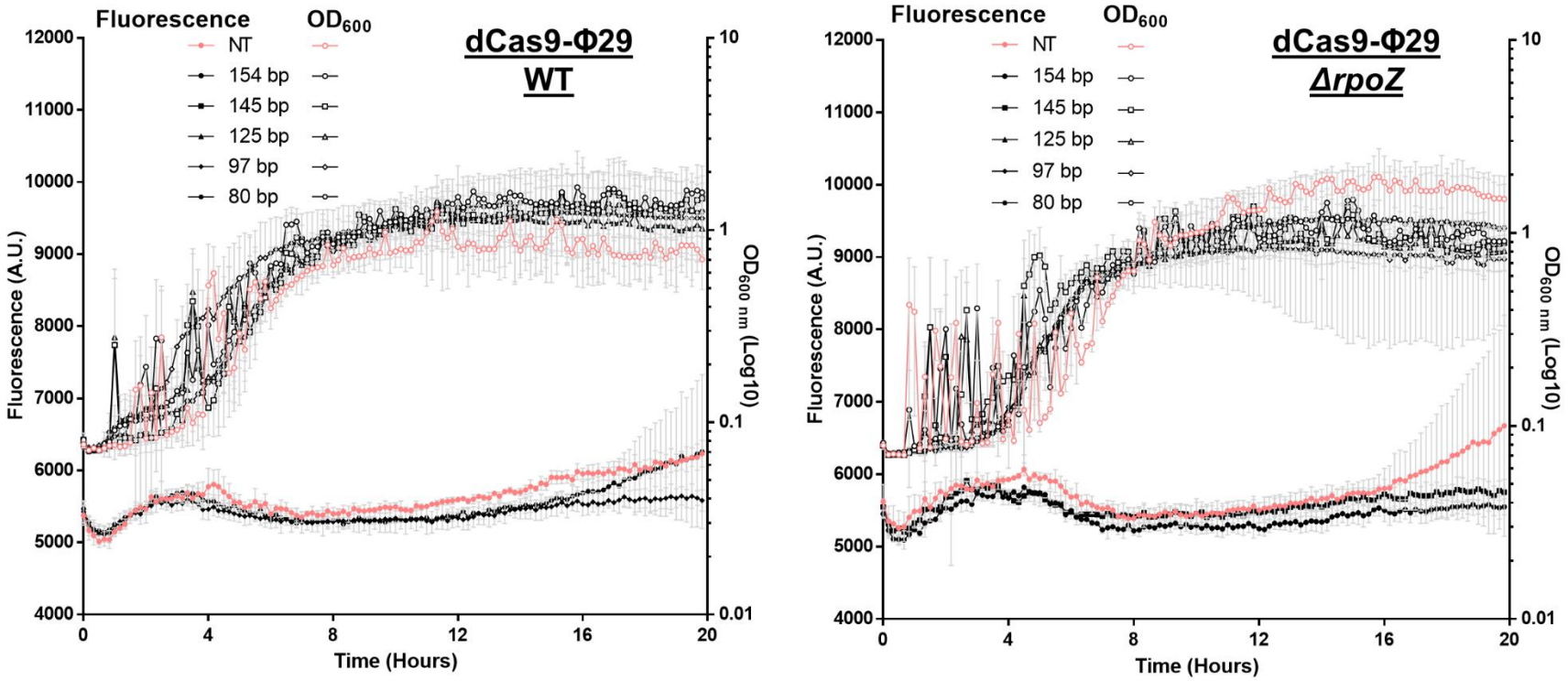


Figure 5.6 - Screening of a dCas9-Φ29 enabled CRISPRa system based on GFPmut3 transcriptional reporters.

Levels of fluorescence (solid symbols) and OD $_{600\ nm}$ (hollow symbols). *gfpmut3* expression is controlled by P_{liaG}. In either the 'WT' BAC0205 (left) or a *ΔrpoZ* BAC0206 (right) background strains, a plasmid is present expressing the candidate CRISPRa dCas9-effector fusion protein, dCas9-Φ29. Each plasmid expressed an sgRNA targeting PAM sites at the indicated distance from the 5´ of the -35 element of P_{liaG}. A non-targeting (NT) sgRNA control strain for each background strain was included (pink). All screened sgRNA targeted PAM sites on the sense strand. Error bars indicate the standard deviation between three biological replicates.

These findings informed the selection of the dCas9-ω(Bs) fusion protein in the BAC0205 background strain for further analysis using the RoboLector micro-fermentation system. Furthermore, the BAC0288 and BAC0289 background strains (where *gfpmut3* is under the control of P_{veg} and $P_{rrnB\ P1}$ respectively) were also included in subsequent analysis to determine if CRISPRa was detectable when the promoter of the targeted gene is of a greater strength. Additional sgRNAs were designed and the subsequent plasmids were cloned to target dCas9-ω(Bs) a range of sites upstream or within the -35 element for each promoter (P_{liaG}, 11 to 97 bp; $P_{rrnB\ P1}$, -20 to 200 bp; P_{veg}, 16 to 200 bp).

Figure 5.7, Figure 5.8 and Figure 5.9 show the RoboLector analysis (section 2.2.3) of all the tested strains, relative to the non-targeting control strain, for the three transcription reporters with weak (P_{liaG} strains BAC243-BAC0245 and BAC0279-BAC0286), intermediate ($P_{rrnB\ P1}$ strains BAC0305-BAC0315) and strong (P_{veg}, strains BAC0295-BAC0304) promoters (Figure 5.3). Tuneable transcription activation could not be detected with the GFPmut3 fluorescence reporters. However, an upregulation in fluorescence, relative to the non-targeting control strain, was noted from 18 hours onwards when dCas9-ω(Bs) was targeted 30 nt upstream of the P_{liaG} -35 element (strain BAC0284), with the highest GFPmut3 fluorescence detected after 24 hours of growth (Figure 5.7). Additionally, targeting dCas9-ω(Bs) 163 nt upstream of the P_{veg} -35 element (strain BAC0303), an increase in relative fluorescence was detected from 3 hours to 4.5 hours growth (Figure 5.9). The same plasmid did not elicit the same response in the equivalent $P_{rrnB\ P1}$ strain (BAC0312 (Figure 5.8)).

sgRNAs targeting PAM sites in close proximity to, or -20 nt downstream, of the -35 element were included as controls to show a decrease in *gfpmut3* expression levels by CRISPRi of up to 85.5% (Figure 5.9), thus confirming the ability of the fusion protein to bind the DNA as designed. RoboLector analysis indicated these controls act as expected with the expression of dCas9-ω(Bs) lowering detectable GFPmut3 fluorescence when strains utilising $P_{rrnB\ P1}$ (Figure 5.8) or P_{veg} (Figure 5.9) promoters were used. However, no lowering of expression levels was observed in P_{liaG} background strains (Figure 5.7). This was thought to be due to the very low expression levels observed (see Figure 5.3) making any reduction in expression levels below the limit of detection.

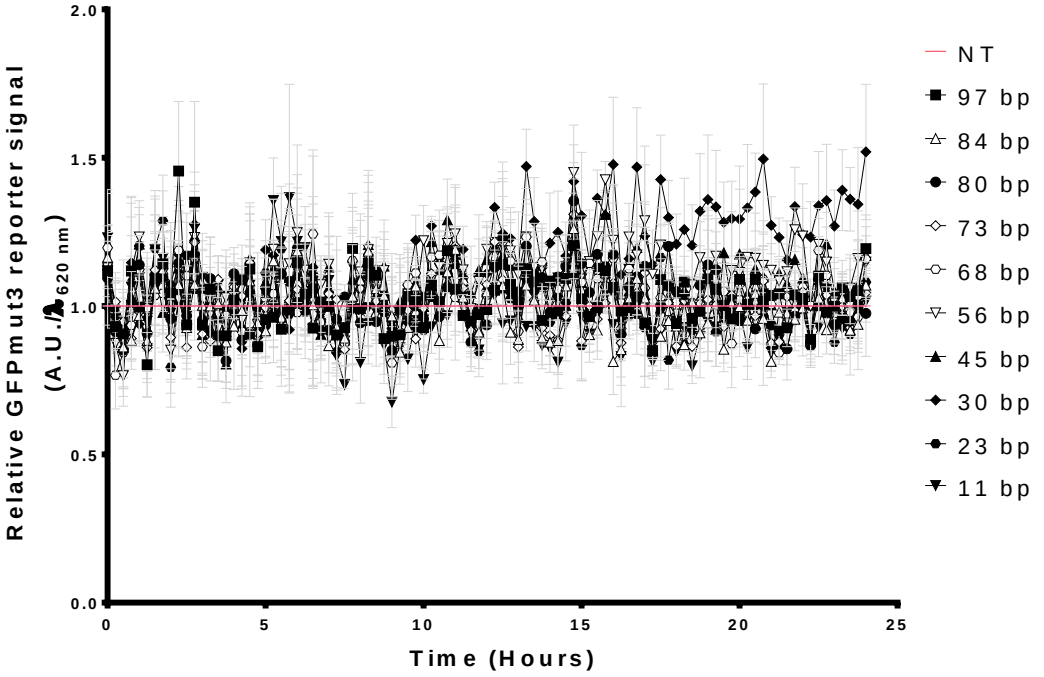


Figure 5.7 – CRISPRa screen of GFPmut3 transcriptional reporters expressed by P_{liaG}.

Levels of GFPmut3 fluorescence in strains with plasmids expressing dcas9-ω(Bs) and sgRNA targeting PAM sites at the indicated distance from the 5´ of the -35 element of the P_{liaG} promoter. Fluorescence levels are normalised by biomass ($\lambda_{620\ nm}$) relative to the non-targeting (NT) sgRNA control strain for each background promoter. sgRNAs target PAM sites on either sense (solid shapes) or antisense strand (hollow shapes). Error bars indicate the standard deviation between three biological replicates.

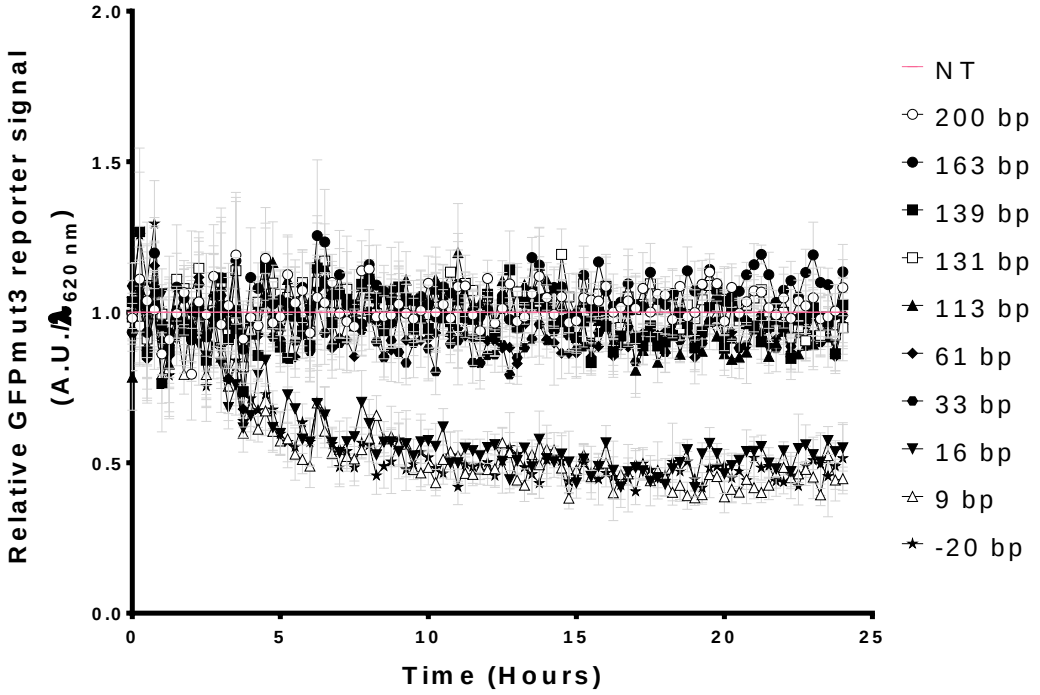


Figure 5.8 – CRISPRa screen of GFPmut3 transcriptional reporters expressed by $P_{rrnB\ P1}$.

Levels of GFPmut3 fluorescence in strains with plasmids expressing dcas9-ω(Bs) and sgRNA targeting PAM sites at the indicated distance from the 5´ of the -35 element of the $P_{rrnB\ P1}$ promoter. Fluorescence levels are normalised by biomass ($\lambda_{620\ nm}$) relative to the non-targeting (NT) sgRNA control strain for each background promoter. sgRNAs target PAM sites on either sense (solid shapes) or antisense strand (hollow shapes). Error bars indicate the standard deviation between three biological replicates.

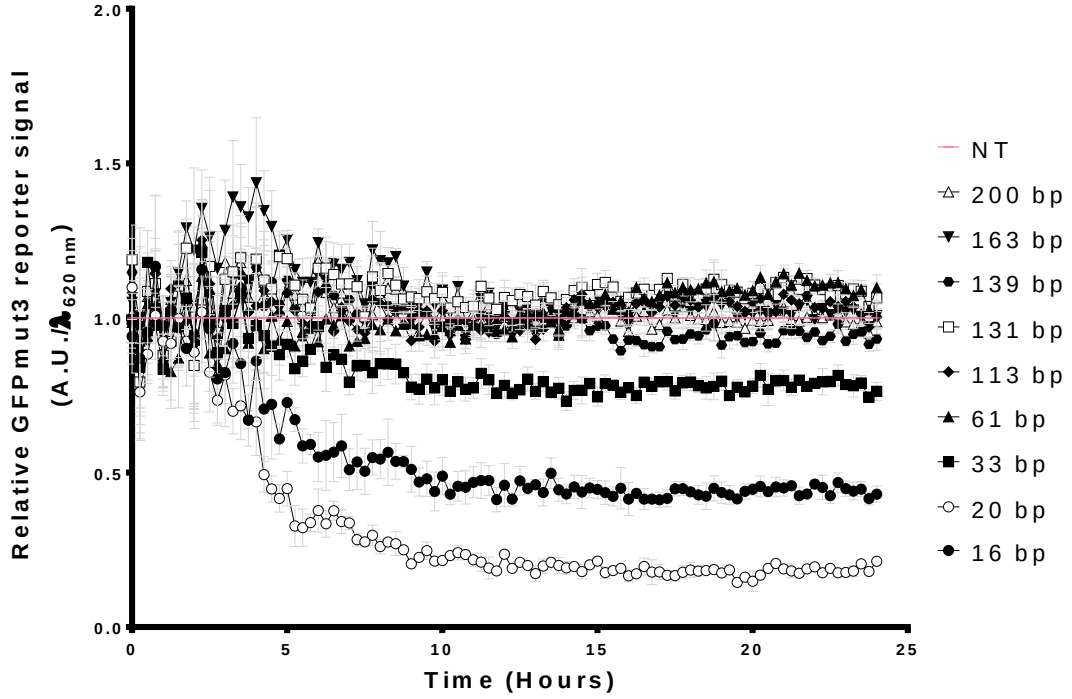


Figure 5.9 – CRISPRa screen of GFPmut3 transcriptional reporters expressed by P_{veg}.

Levels of GFPmut3 fluorescence in strains with plasmids expressing dcas9-ω(Bs) and sgRNA targeting PAM sites at the indicated distance from the 5´ of the -35 element of the P_{veg} promoter. Fluorescence levels are normalised by biomass ($\lambda_{620\ nm}$) relative to the non-targeting (NT) sgRNA control strain for each background promoter. sgRNAs target PAM sites on either sense (solid shapes) or antisense strand (hollow shapes). Error bars indicate the standard deviation between three biological replicates.

Taking the potential CRISPRa observations and CRISPRi control strains, Figure 5.10, Figure 5.11 and Figure 5.12 investigates the relationship between fluorescence and distance of the PAM or protospacer to the -35 element. The timepoint for each background strain yielding the highest fluorescence was analysed (P_{liaG} – 24 hours; $P_{rrnB\ P1}$ – 6.25 hours; P_{veg} – 4 hours). Using regression analysis, an R^2 value for each analysis was established. This showed a closer relationship to the distance of the distal end of the protospacer to the level of fluorescence observed for both the $P_{rrnB\ P1}$ and P_{veg} strain sets. However, no correlation could be observed for the P_{liaG} strains given there was no significant difference in fluorescence.

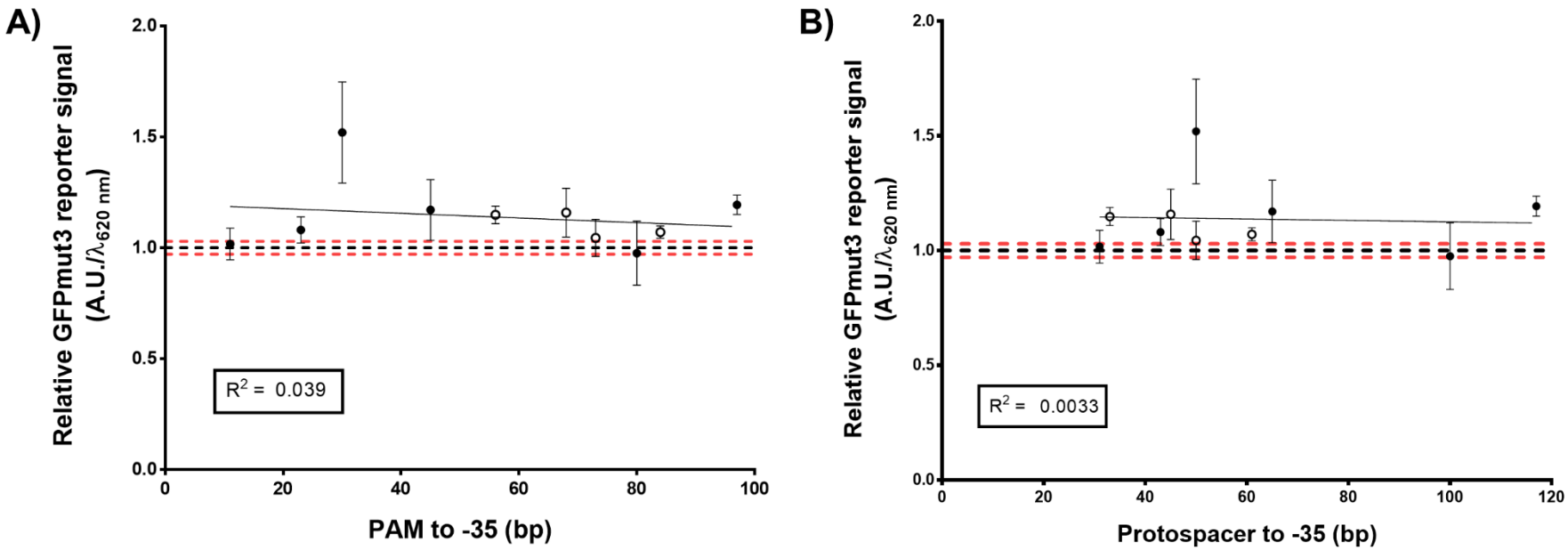


Figure 5.10 – Correlation between P_{liaG} controlled GFPmut3 relative fluorescence and distance from PAM site (A) or protospacer (B) to the -35 element.

Strains with *gfpmut3* under the control of P_{liaG} at 24 hours growth. Black dashed lines represent the non-targeting sgRNA control, with red dashed lines representing the corresponding standard deviation between three biological replicates. A solid black line of best fit is shown for each graph, and the corresponding R^2 value is also shown. Error bars represent the standard deviation between three biological replicates.

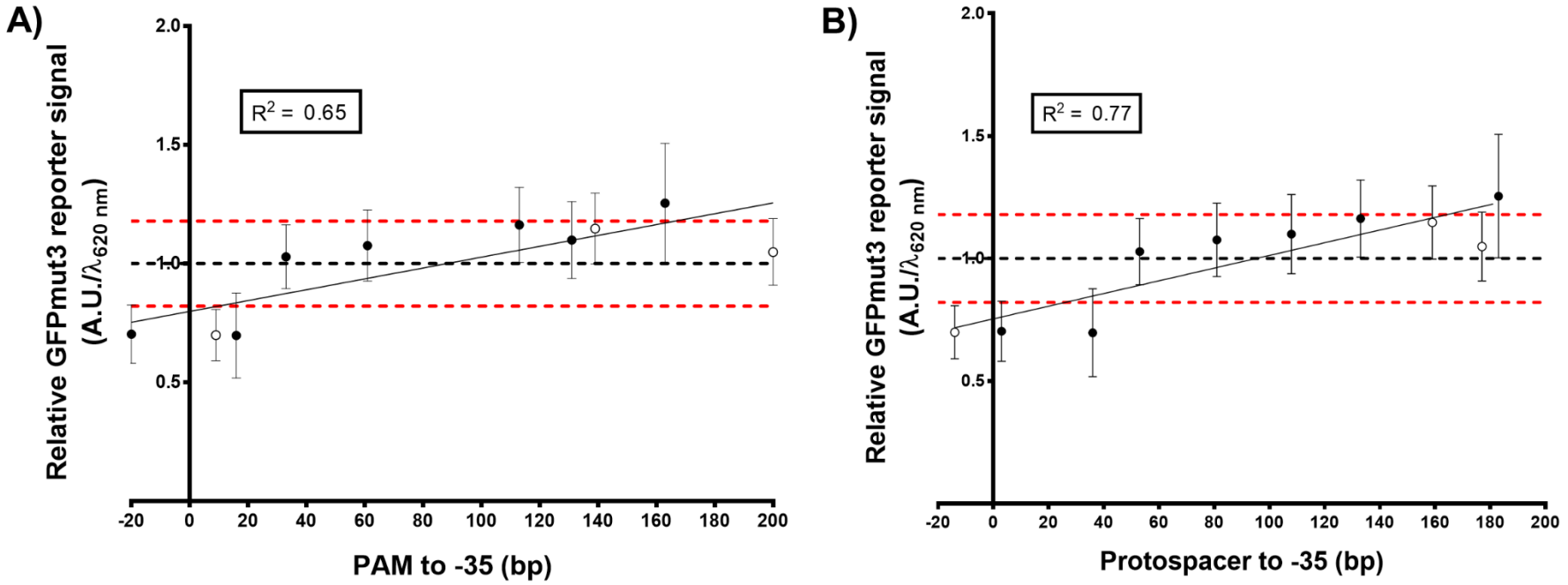


Figure 5.11 – Correlation between $P_{rrnB\ P1}$ controlled GFPmut3 relative fluorescence and distance from PAM site (A) or protospacer (B) to the -35 element.

Strains with *gfpmut3* under the control of $P_{rrnB\ P1}$ at 24 hours growth. Black dashed lines represent the non-targeting sgRNA control, with red dashed lines representing the corresponding standard deviation between three biological replicates. A solid black line of best fit is shown for each graph, and the corresponding R^2 value is also shown. Error bars represent the standard deviation between three biological replicates.

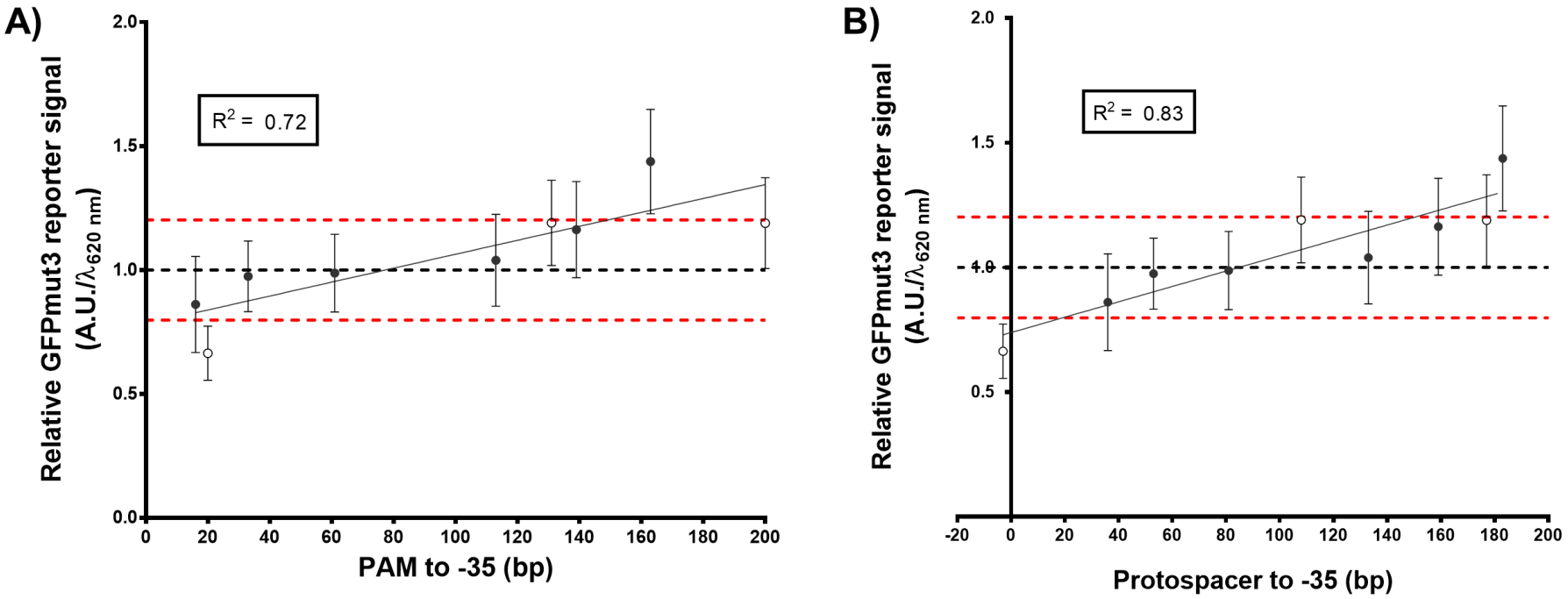


Figure 5.12 – Correlation between P_{veg} controlled GFPmut3 relative fluorescence and distance from PAM site (A) or protospacer (B) to the -35 element.

Strains with *gfpmut3* under the control of P_{veg} at 24 hours growth. Black dashed lines represent the non-targeting sgRNA control, with red dashed lines representing the corresponding standard deviation between three biological replicates. A solid black line of best fit is shown for each graph, and the corresponding R^2 value is also shown. Error bars represent the standard deviation between three biological replicates.

5.3.1.3 Analysis of CRISPRa effect by quantification of *gfpmut3* mRNA confirms transcriptional interference but not activation

To validate if the potential CRISPRa we had observed was a true reflection of an increase in the mRNA levels, RT qRT-PCR analysis was performed (section 2.14). Since the strains where *gfpmut3* was expressed using P_{veg} had given the greatest correlation between transcriptional changes, relative to the non-targeting strain, and the distances of either the PAM site, or protospacer from the -35 promoter element (Figure 5.12), these were selected for analysis. Strain BAC0303 was selected for CRISPRa analysis since a 43.7% increased level in GFPmut3 reporter had been detected during exponential growth phase (Figure 5.9). As a control, the P_{veg} strain BAC0295 was analysed to show the decrease in relative fluorescence observed was due to dCas9-ω(Bs) inhibiting transcription.

5.3.1.4 Validation of RT qRT-PCR primers

Two primer sets were designed and validated (section 2.14.5) for *gfpmut3* (Figure 5.13), as well as the *gapA* (Figure 5.14) and *sdhA* (Figure 5.15) reference genes which are constitutively expressed under the experimental conditions in *B. subtilis*[166,167]. As adequate RT qPCR primers are described as allowing a PCR efficiency of 90-110%, all the tested primer sets were adequate for use. The primer sets for each gene yielding a PCR efficiency closest to 100% were selected for use in subsequent strain analysis (*gfpmut3* set 1, *gapA* set 2, *sdhA* set 1).

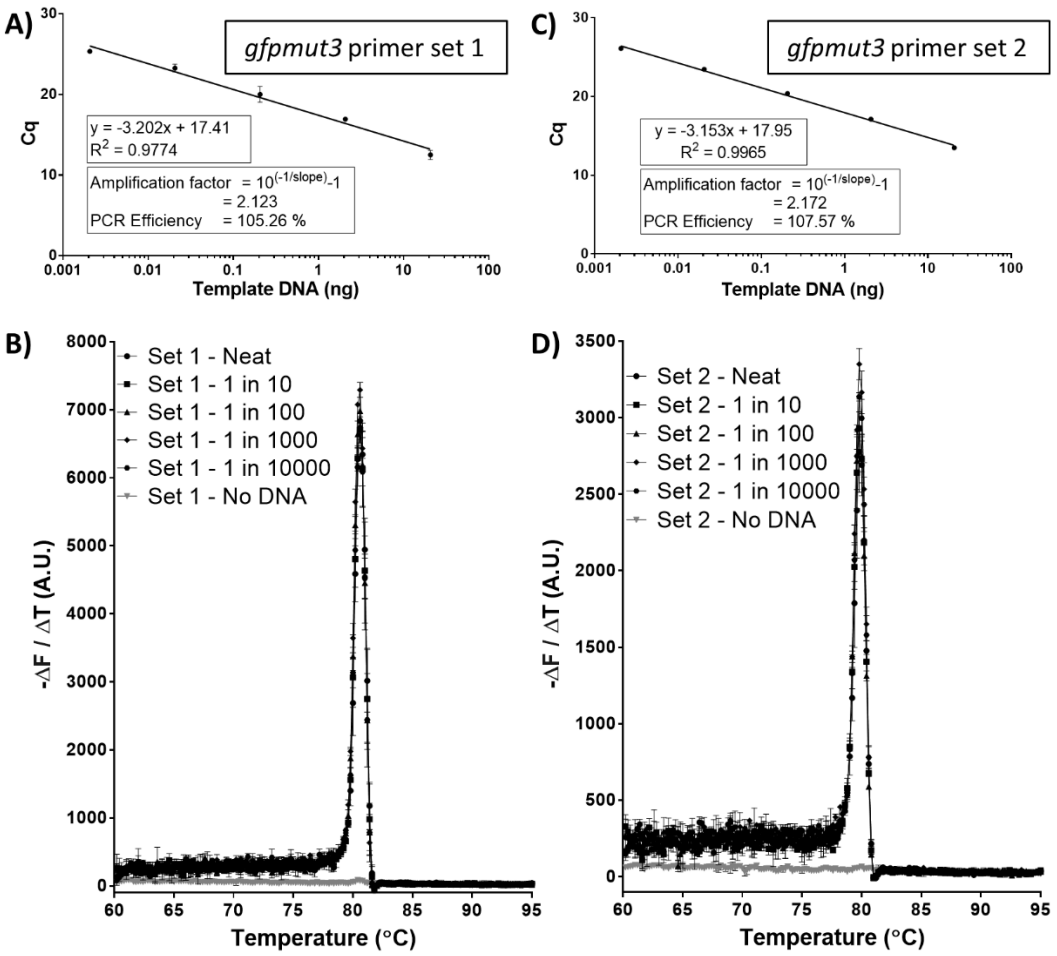


Figure 5.13 – Validation of RT qRT-PCR primer pair for analysis of *gfpmut3*.

Cq (quantification cycle) calibration curve with the subsequent PCR efficiency calculation (**A** and **C**), and melt curve results (**B** and **D**) for each primer set. Melt curves were carried out following RT qPCR with serial dilutions of BAC0288 gDNA, with a no DNA control reaction also present (grey line). *gfpmut3* primer set 1 was selected for use in analysis of *gfpmut3* transcription.

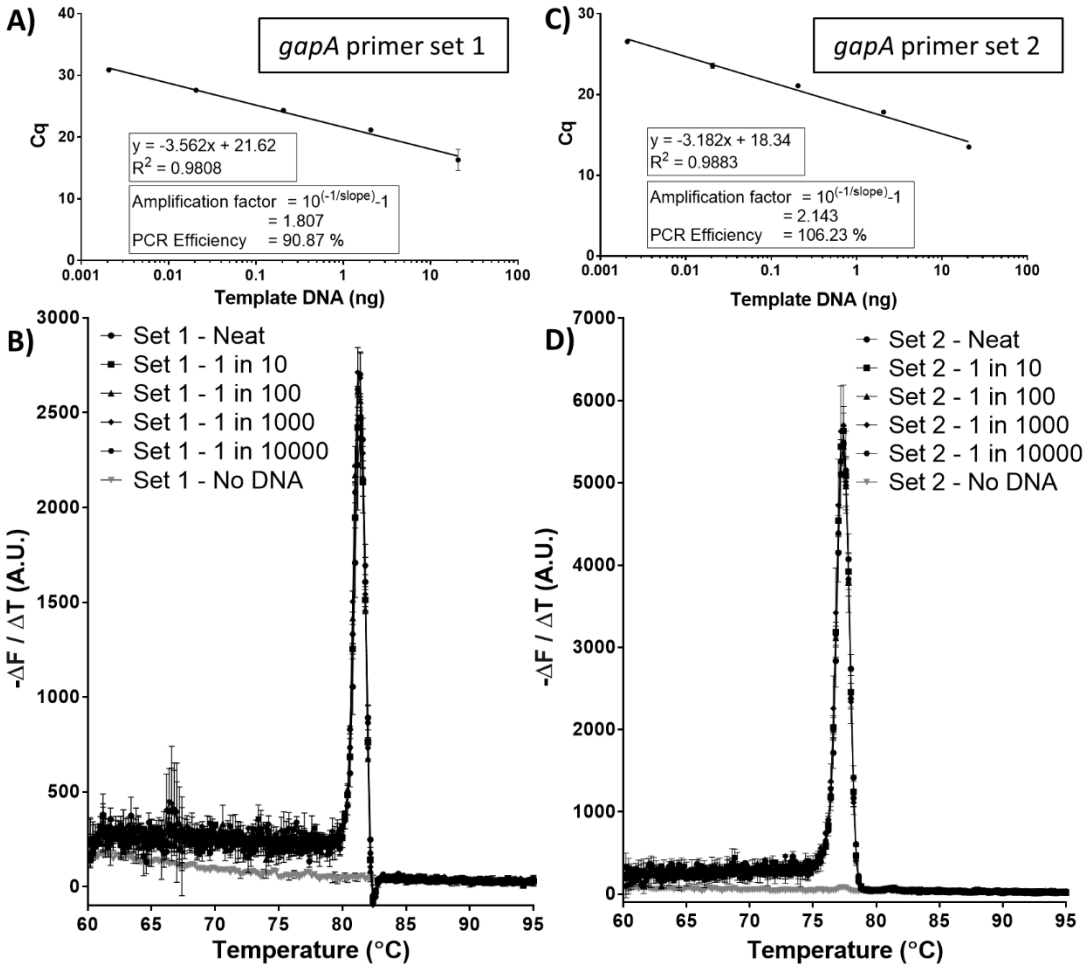


Figure 5.14 – Validation of RT qRT-PCR primer pair for analysis of *gapA*.

Cq (quantification cycle) calibration curve with the subsequent PCR efficiency calculation (**A** and **C**), and melt curve results (**B** and **D**) for each primer set. Melt curves were carried out following RT qPCR with serial dilutions of BAC0288 gDNA, with a no DNA control reaction also present (grey line). *gapA* primer set 2 was selected for use in analysis of *gapA transcription*.

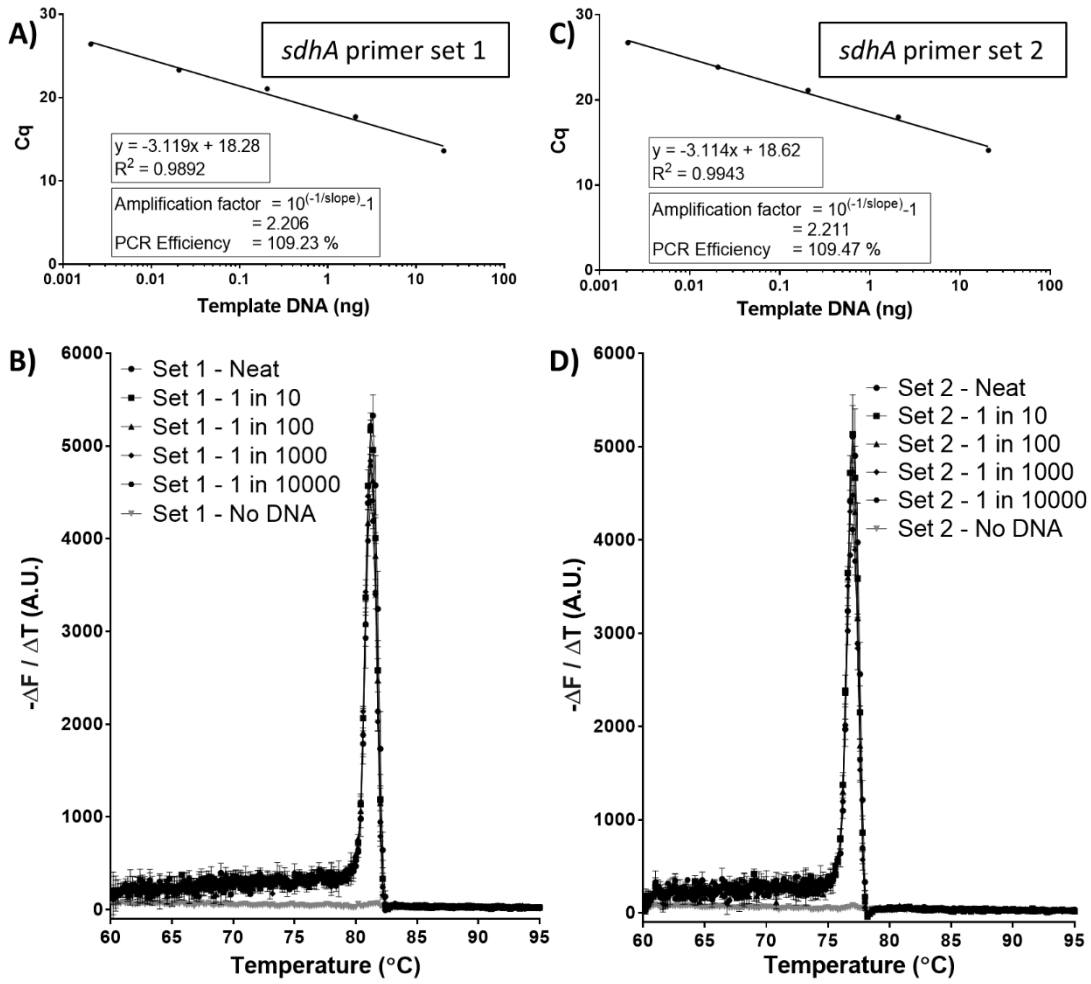


Figure 5.15 – Validation of RT qRT-PCR primer pair for analysis of *sdhA*.

Cq (quantification cycle) calibration curve with the subsequent PCR efficiency calculation (**A** and **C**), and melt curve results (**B** and **D**) for each primer set. Melt curves were carried out following RT qPCR with serial dilutions of BAC0288 gDNA, with a no DNA control reaction also present (grey line). sdhA primer set 1 was selected for use in analysis of *sdhA* transcription

5.3.1.5 RT qRT-PCR confirms *gfpmut3* downregulation in control strain BAC0295 and lack of transcriptional activation in assay strain BAC0303

RT qRT-PCR was used to quantify *gfpmut3* transcription in strains BAC0295 and BAC0303 (section 2.14). RT qRT-PCR results can be found in Appendix F. These transcription levels were compared to those of strain BAC0297 which contained the non-targeting control plasmid pBAC0104, as well as the reference

genes *gapA* and *sdhA* (Figure 5.16). The results confirm the lowered fluorescence observed in BAC0295 was due to a lower mRNA pool caused by dCas9-ω(Bs) blocking efficient binding of the RNAP to the P_{veg} promoter. There was, however, no statistically significant increase in *gfpmut3* levels in BAC0303. This confirmed that CRISPRa was not observed under the tested conditions.

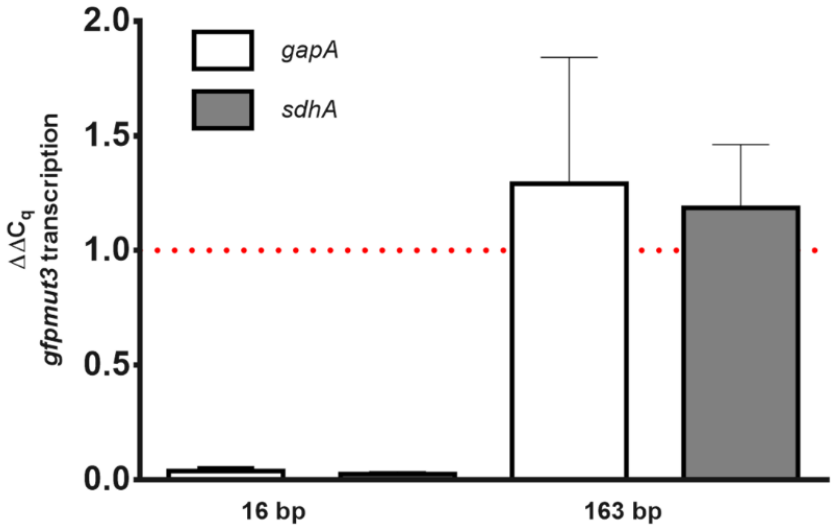


Figure 5.16 – Relative *gfpmut3* expression following CRISPR-mediated transcriptional regulation.

gfpmut3 transcription levels relative to *gapA* and *sdhA* transcription levels between assay strains BAC0295 (16 bp (targeting PAM site 16 bp upstream of the -35 region of P_{veg})) or BAC0303 (163 bp (targeting PAM site 163 bp upstream of the -35 region of P_{veg})) and the non-targeting control strain BAC0297 ($\Delta\Delta C_q$). $\Delta\Delta C_q$ to *gapA* and *sdhA* was calculated as per Equation 2-5 and Equation 2-6 respectively. RT qRT-PCR results can be found in Appendix F. Error bars indicate the standard deviation between technical duplicates and biological triplicates.

5.3.2 Multiplexed CRISPRi with dCas9

The control reactions in section 5.3.1 exemplified CRISPRi for the first time using the plasmid system described in this study (Figure 5.8 and Figure 5.9). It was desirable to determine if this single plasmid system also allowed the simultaneous targeting of dCas9 to more than one locus (multiplexing). BAC0288 was selected as the test strain for this analysis as it had been shown to be an effective reporter for *gfpmut3* expression downregulation (Figure 5.9). Furthermore, the native *amyE*

gene, encoding a starch degrading α-Amylase, was also used as a second reporter protein for CRISPRi.

Multiplexed targeting requires more than one gRNA. Since the expression of each sgRNA requires an individual expression cassette (promoter-sgRNA-terminator), it was decided to simplify the design by using the native gRNA model from *S. pyogenes*[43]. The sgRNA was therefore divided again into its constitutive tracrRNA and crRNA components, with the new parental multiplexing CRISPRi, *dcas9* expressing plasmid, pBAC0164, constructed to contain a tracrRNA expression cassette, controlled by $P_{rrnB\ P1}$, and a crRNA array expression cassette, controlled by P_{veg} (Figure 5.1).

The crRNA array was designed with three protospacers inserted targeting, in order of expression, *amyE*, *gfpmut3* and a non-targeting protospacer. This was cloned into pBAC0164 by a three-part ligation of the pBAC0164 SapI digested backbone and two sets of annealed oligonucleotides with a 4 nt complimentary overhang to each other and 3 nt to the backbone. The joining site between the two inserted DNA oligonucleotides was the direct repeat region which had been split to avoid repeat sequences which are known to complicate DNA synthesis. This region can be further split (as described in section 2.17) with up to 12 complementary overhangs available (GC content: 25-75% and at least 2 bp difference to limit non-specific binding). Thus, where two protospacers are present on the annealed oligonucleotide, either side of a direct repeat region, and the subsequent linking region is present, up to 24 protospacers could be cloned within a single step.

Strains BAC0331, expressing *dcas9*, tracrRNA and the crRNA array, and BAC0330, expressing *dcas9* and tracrRNA only, were grown and analysed following 4 and 20 hours of growth for extracellular α-Amylase activity (section 2.9.2), and intracellular GFPmut3 fluorescence levels (section 2.10.2).

The results (Figure 5.17), relative to the BAC0330 negative control strain without crRNA, show that CRISPRi was effective for downregulation of both target reporter proteins. The α-Amylase activity results following 4 hours of growth show a broad variation across the analysed replicates (58.7% ± 38%). This was thought to be due to expression of *amyE* occurring primarily in the stationary phase of growth and as such the level of α-Amylase in the extracellular environment is too low to accurately quantify due to the assay sensitivity being too low. The α-Amylase

activity levels observed following 20 hours of growth show an average, across the biological replicates, of 52.7% reduction when compared to the control strain.

Since GPFmut3 is retained within the cytoplasm, and it is constitutively expressed by P_{veg} (Figure 5.3), efficient downregulation by CRISPRi could be seen during exponential growth at 4 hours, with an average of 75.1% reduction compared to the control strains (Figure 5.17). Following 20 hours of growth, a similar CRISPRi effect of 71.8% was observed. This was an increase in CRISPRi efficiency when compared to BAC0298 (BAC0288 with dCas9-ω(Bs) and the same targeted PAM site) where a knock-down efficiency of 55.4% was achieved following 20 hours of growth (Figure 5.9).

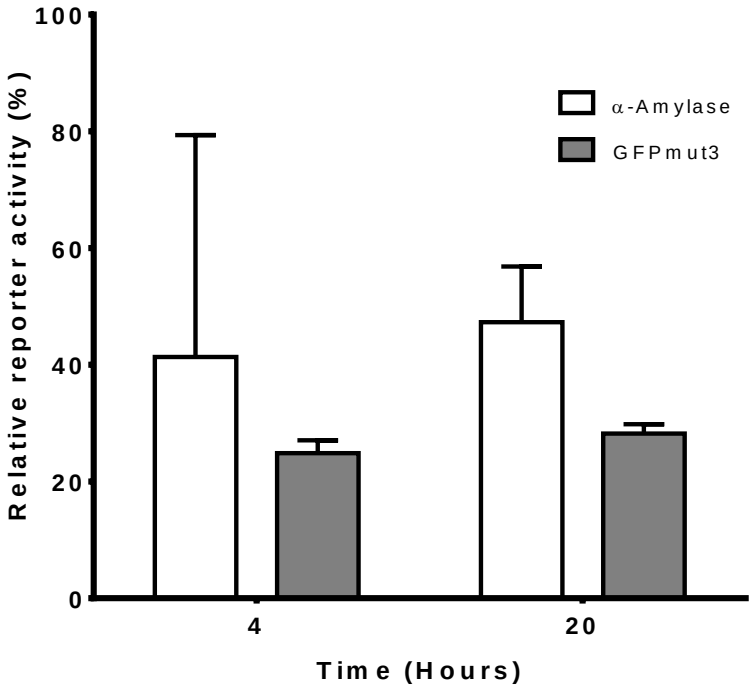


Figure 5.17 – Relative α-Amylase activity and GFPmut3 fluorescence following dCas9 mediated multiplexed CRISPRi.

Bars represent extracellular α-Amylase activity (section 2.9.2), or GFPmut3 fluorescence intensity (section 2.10.2) normalised by $OD_{600\ nm}$ for BAC0331 relative to a control strain (BAC0330). Activity was measured at 4 and 20 hours of growth. Error bars indicate the average standard deviation between three technical replicates for three biological replicates.

5.3.3 Engineering and characterisation of dMAD7

The alternative CRISPR nuclease, MAD7, has been shown as functional for CRISPR enabled genome editing in *E. coli*, *S. cerevisiae,* mouse and human HEK293T cells[67], zebrafish[71], and *B. subtilis* in this study (section 3.3.2). Due to MAD7s status as free to use for both academic and commercial applications[70], we decided to engineer a catalytically inactive variant of MAD7 to allow licence free CRISPRi.

5.3.3.1 Identification of MAD7 catalytic residues

A pairwise alignment of the amino acid sequences of MAD7 and AsCpf1, confirmed the 31% identity reported by Inscripta (Figure 5.18). Previously, it was reported that the catalytic residues of AsCpf1 are Asp908, Glu993 and Asp1263[68,137]. Asp908 lies in a region of high similarity with MAD7, with residues 905-916 corresponding to MAD7 residues 874-885. AsCpf1 Glu993 does not lie in a region of high homology, however the alignment revealed that this residue was conserved in MAD7 (Figure 5.18). Lastly, the residue corresponding to Asp1263 in AsCpf1 was found in a region of high homology with AsCpf1, with residues 1261-1268 corresponding to MAD7 residues 1211-1218. The corresponding catalytic residues in MAD7 (Asp877, Glu962 and Asp1213) identified by sequence homology (Figure 5.18) were simultaneously modified to alanine *in silico* and the corresponding gene *dMAD7* was synthesised.

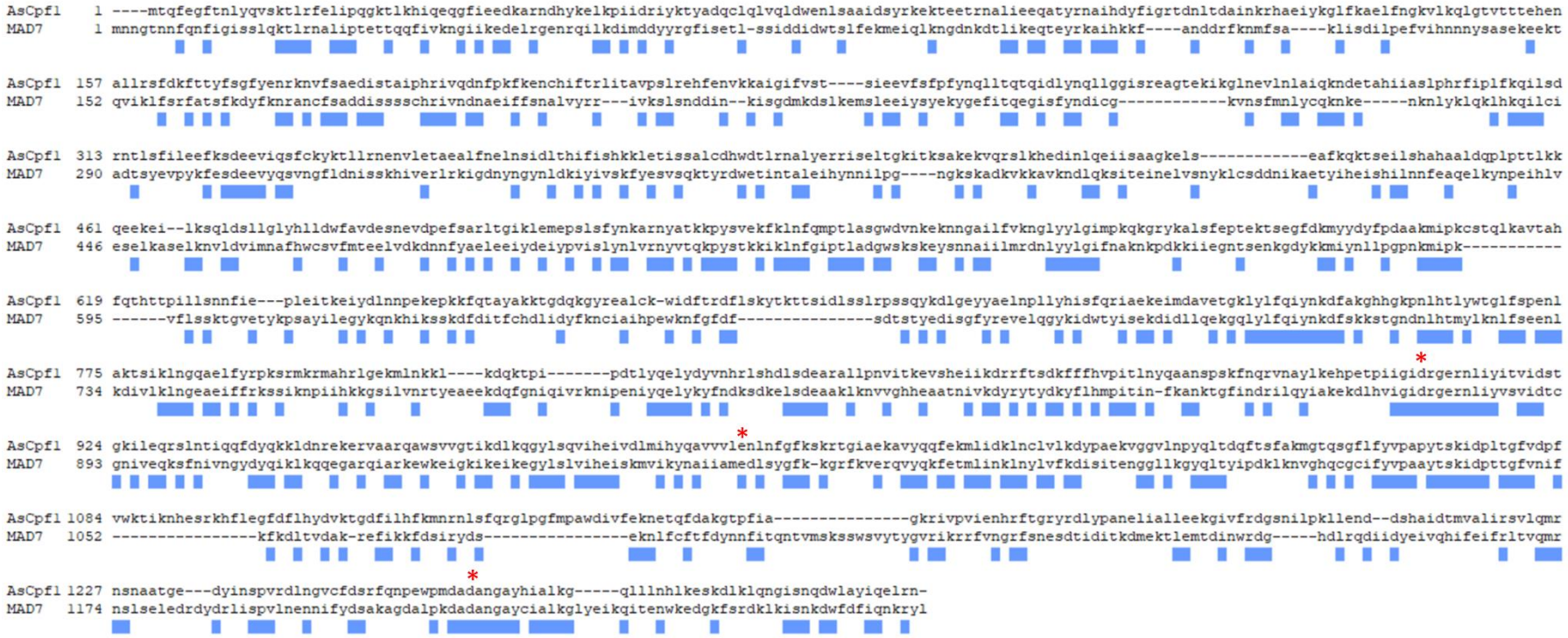


Figure 5.18 – Pairwise sequence alignment of AsCpf1 and MAD7 amino acid sequence for identification of the catalytic residues in MAD7.

Blue bars represent residues of similarity between the AsCpf1 and MAD7. Red asterixis above the sequence represent the three catalytic residues identified in MAD7 - Asp877, Glu962 and Asp1213.

5.3.3.2 Absence of nuclease activity

To verify whether the DNA cleavage capacity of MAD7 was removed in the putative dMAD7, the synthesised *dMAD7* gene was used to replace *MAD7* within the *amyE* and *gfpmut3* targeting plasmids pBAC0162 and pBAC0166, respectively. As a control, the well characterised *dcas9* was used to replace *cas9* in plasmids pBAC0041 and pBAC0165[73,135,138]. Naturally competent *B. subtilis* 168 and BAC0288 were respectively transformed with the *amyE* and *gfpmut3* targeting plasmids with both active and inactive nuclease variants. Transformations were spread on plates supplemented with chloramphenicol and IPTG to ensure nuclease expression. The number of CFUs obtained for each transformation (Table 5.2) indicates that the engineered dMAD7 does not catalyse DSB of DNA since it does not cause the reduced viability observed for the catalytically active nuclease.

Table 5.2 – Average number of transformants obtained following three transformations of *B. subtilis* 168 and BAC0288 with Cas9, dCas9, MAD7 and dMAD7 plasmids.

Parental Strain	Target	Plasmid	Nuclease	Average CFU obtained
***B. subtilis* 168**	*amyE*	pBAC0041	Cas9	0.3
		pBAC0184	dCas9	898
		pBAC0162	MAD7	1.7
		pBAC0188	dMAD7	1248
BAC0288	*gfpmut3*	pBAC0165	Cas9	0
		pBAC0195	dCas9	392
		pBAC0166	MAD7	0.7
		pBAC0190	dMAD7	343

5.3.3.3 Retention of DNA binding capacity for CRISPRi

Extracellular α-Amylase activity was quantified in strains expressing dMAD7 targeting five PAM sites (5´-TTTN-3') at the 5´ end of *amyE*, two on the template strand and three on the non-template strand. The results were directly compared to strains expressing dCas9 targeting *amyE*. Strains expressing non-targeting dMAD7

and dCas9 plasmids were used as the negative controls for downregulation (Figure 5.19B).

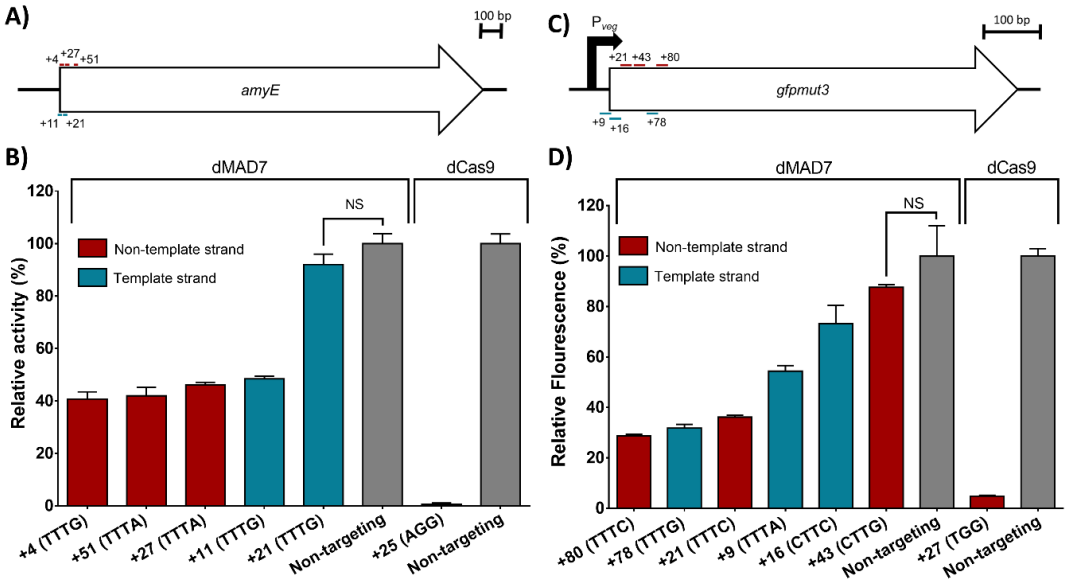


Figure 5.19 – Relative α-Amylase activity and GFPmut3 fluorescence following dCas9 and dMAD7 mediated CRISPRi.

A) and **C)** Schematic diagrams of gRNA binding sites for dMAD7 within *amyE* and *gfpmut3* respectively. Values represent the distance of each targeted PAM site from the start codon on either the template (blue) or non-template (red) DNA strand. **B)** Bar graph represents the extracellular α-Amylase activity *(*section 2.9.2*)* normalised by $OD_{600\ nm}$ relative to the non-targeting gRNA control for either dMAD7 or dCas9, after 24 hours of growth. The horizontal axis indicates the PAM site targeted by each gRNA and its distance to the *amyE* start codon (+4: BAC0360; +51: BAC0363; +27: BAC0362; +11: BAC0361; +21: BAC0355; +25: BAC0353). Red and blue bars correspond to PAM sites on the non-template and template strands, respectively. Grey bars represent non-targeting controls (dMAD7: BAC0354; dCas9: BAC0352). **D)** Bar graph represents the fluorescence intensity *(*section 2.10.2*)* normalised by $OD_{600\ nm}$ relative to the non-targeting gRNA control for either dMAD7 or dCas9, after 24 hours of growth. The horizontal axis indicates the PAM site targeted by each gRNA and its distance to the *gfpmut3* start codon (+80: BAC0368; +78: BAC0367; +21: BAC0351; +9: BAC0364; +16: BAC0365; +43: BAC0366; +27: BAC0349). Red and blue bars correspond to PAM sites on the non-template and template strands, respectively. Grey bars represent the non-targeting controls (dMAD7: BAC0350; dCas9: BAC0348). Error bars indicate the standard deviation between three biological replicates. NS = not statistically significant two-tailed P values following unpaired t test with Welch's correction.

The results confirm that dCas9-mediated CRISPRi is highly efficient with a 99.4% reduction in α-Amylase activity, while dMAD7-mediated CRISPRi appears less efficient, ranging from 59.3% to 51.5% activity reduction depending on the gRNA and PAM site sequence. The gRNA targeting the PAM site TTTG +21 bases from the start codon did not exhibit significantly reduced levels of activity. The PAM sequence was the same as others where CRISPRi was successful, and the GC% of the protospacer (28.6%) is similar to the one targeting the PAM site +4 (33.3%).

To further investigate the capacity of dMAD7 for CRISPRi, the *gfpmut3* gene in strain BAC0288 was targeted. Here, six PAM sites (three on each strand) were targeted with the 5´-YTTN-3' PAM sequence recommended by Inscripta (Figure 5.19C). dMAD7 was targeted to the 5´ end of *gfpmut3* and fluorescence was compared to strains expressing dCas9 targeting *gfpmut3*. Strains expressing non-targeting dMAD7 and dCas9 plasmids were used as negative controls for downregulation (Figure 5.19D). Here, a broader effect on expression was observed for dMAD7 with statistically significant CRISPRi efficiencies ranging from 71.3% - 26.8%. Once again, the *gfpmut3* targeting dCas9 control exhibited highly efficient CRISPRi with a 95.1% reduction in expression.

5.3.3.4 Multiplexing CRISPRi with dMAD7

To increase the efficiency of dMAD7 transcriptional downregulation, multiplexed targeting, where more than one gRNA is utilised at a time, was tested. gRNA arrays were inserted in the same manner as single gRNAs, with the final array designed with the dMAD7 handle direct repeat at both the 3' and 5´ ends, as well as between the inserted protospacers.

Additive downregulation of *amyE* was tested by targeting PAM sites at +4 TTTG and +51 TTTA in strain BAC0377 (Figure 5.20B). Similarly, *gfpmut3* was analysed by targeting PAM sites at +80 TTTC and +21 TTTG in strain BAC0380 (Figure 5.20B). These were compared to the non-targeting dMAD7 control strain BAC0350, as well as the single gRNA, *amyE* or *gfpmut3* targeting dMAD7 strains BAC0381, BAC0382, BAC0351 and BAC0368. To ensure the gRNA array was matured from a single transcript into single gRNA units, strain BAC0378, carrying a

gRNA array to target *amyE* +4 TTTG and *gfpmut3* +80 TTTC, was analysed for transcriptional interference of both targets (Figure 5.20C).

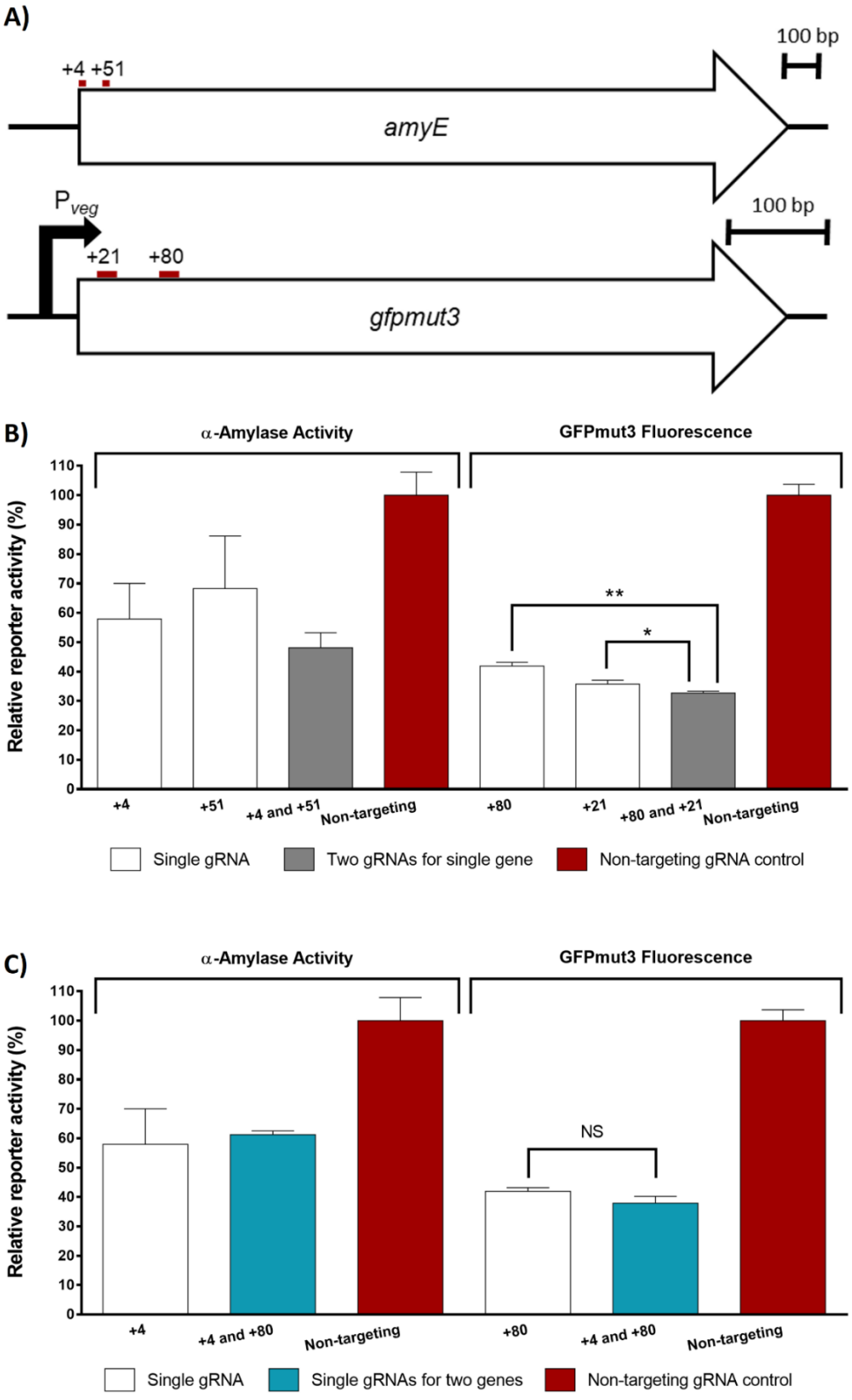


Figure 5.20 – Relative α-Amylase activity and GFPmut3 fluorescence following dMAD7 mediated multiplexing CRISPRi in BAC0288 (previous page).

A) Schematic diagrams of gRNA binding sites for dMAD7 within *amyE* and *gfpmut3*. Values represent the distance of each targeted PAM site from the start codon on the non-template DNA strand. **B**) Investigation for an additive multiplexed CRISPRi effect when two gRNAs target a single gene. **C**) Investigation for a multiplexed CRISPRi effect when two gRNAs target different genes within the same strain. Bars represent extracellular α-Amylase activity (section 2.9.2), or GFPmut3 fluorescence intensity (section 2.10.2), normalised by $OD_{600\ nm}$, relative to the non-targeting gRNA control strain, after 24 hours of growth. The horizontal axis indicates the targeted PAM site(s) distance to the *amyE* or *gfpmut3* start codon (Strains used: +4: BAC0381; +51: BAC0382; +4 and +51: BAC0377; +4 and +80: BAC0378; +80: BAC0368; +21: BAC0351; +80 and +21: BAC0380). White bars indicate strains in which a single gRNA is utilised to target a single gene. Grey bars indicate strains in which the effect of two gRNAs on a single gene is investigated. Blue bars indicate BAC0378 in which the effect of single gRNAs on two different genes is investigated. Red bars indicate the negative control strain, BAC0350, in which a non-targeting gRNA is expressed alongside dMAD7. Error bars indicate the standard deviation between three biological replicates. NS – no significant difference; * – $P < 0.05$; ** – $P < 0.01$. Two-tailed P values were derived following unpaired t test with Welch's correction.

The results show when *amyE* and *gfpmut3* were simultaneously targeted the downregulation of both genes was found to be similar to when only one was targeted for CRISPRi (Figure 5.20C). This confirms that dMAD7, within the *B. subtilis* 168 derivative, BAC0288 background, has the capacity to mature the crRNA array transcript into individual gRNA and there is no significant competition between the two gRNAs for the pool of dMAD7 to mediate CRISPRi. When two gRNAs are combined to target either *gfpmut3* or *amyE,* the measured downregulation is not cumulative (Figure 5.20B).

5.4 Discussion

5.4.1 Cloning efficiency limitations with AarI

During the initial screen of the three dCas9-effector fusion proteins (section 5.3.1.2), some difficulty was encountered in the cloning 7/21 of the protospacer elements to complete the sgRNAs. This was thought to be caused by AarI digestion reactions not going to completion, or only cleaving a single restriction site since the

enzyme requires two copies of the restriction site to initial cleavage. An increased concentration of the oligonucleotide supplied with AarI by Thermo Fisher Scientific (containing an AarI restriction site) from the recommended 1x to 2.5x, and the dephosphorylation of the AarI digested plasmid backbone greatly improved the success of cloning in the subsequent round of transcriptional analysis and was implemented in all gRNA cloning thereafter (section 2.7.3).

5.4.2 CRISPRa was found not to be functional in *B. subtilis*, in contrast to another study by Lu *et al.*

In this chapter it was shown that, under the tested conditions, CRISPRa cannot be achieved in *B. subtilis* 168 when targeting dCas9-ω(Bs) upstream of *gfpmut3*. dCas9-ω(Ec) and dCas9-Φ29 fusion proteins were also tested for enabling a CRISPRa effect of *gfpmut3*, however these yielded no detectable significant increase in relative fluorescence. Bikard *et al*. had shown dCas9-ω(Ec) to act as an effective fusion protein for CRISPRa in *E. coli* were the RNAP ω subunit was deleted (*ΔrpoZ*)[59]. This was based on data where transcription was induced in *E. coli ΔrpoZ* when the ω subunit was fused to the λ bacteriophage repressor protein[163]. As an equivalent study has not been carried out in *B. subtilis*, this would be a prudent next step to ensure the ω subunit can in fact act as a transcriptional activator in this host. A major difference to the approach taken by Bikard *et al*. was that this method targets a chromosomally integrated reporter protein, as opposed to a plasmid-based system. Furthermore, the targeted region upstream of the synthetic promoter used by Bikard *et al.* was fully designed with regularly dispersed PAM sites.

Since performing this study, Lu *et al*. successfully achieved CRISPRa within *B. subtilis* SCK6 and dCas9 fused with the ω or α RNAP subunits[60]. The primary differences that could be discerned from this paper with the study presented here are the *B. subtilis* strain used (SCK6 (Em, *his*, *nprE18*, *aprE3*, *eglSΔ102*, *bglT/bglSΔEV*, *lacA::PxylA-comK*) vs. the strains reported here of a *B. subtilis* 168 (*trpC2*) background), as well as the distance from the transcriptional start site (TSS) to which the dCas9-ω/α was targeted (from 267 to 415 (Lu *et al*.) versus 51 to 235 bases upstream of the TSS). A further significant difference is the linker region connecting the ω or α subunits to dCas9 (Gly-Ser-Ala-Ala-Ser by Lu *et al*. vs. Ala-Ala here, which was based on the linker used by Bikard *et al*.). This increased

linker size could enable greater flexibility for the subunits to better access and attract the RNAP.

5.4.2.1 Alternative transcriptional activator subunits to fuse with dCas9 to enable CRISPRa in *B. subtilis*

Both in this study, and in that carried out by Lu *et al*., the ω subunit was the primary effector molecule studied for the enhancement of transcriptional activation. However, other proteins have been found to be more efficient transcriptional activators in *E. coli*. Dong *et al*. found that one such activator, SoxS, in a WT *E. coli* strain, yielded ~8-fold greater GFP expression compared to the ω subunit, and ~2-fold greater when the ω subunit activator was used in a *ΔrpoZ* background[168]. This strategy employed the use of an extended sgRNA, in which an RNA scaffold protrudes from the Cas9 upon which the bacteriophage coat protein MS2 RNA binding domain, fused to SoxS, could bind to enable site specific RNAP recruitment. CRISPRa was found to be functional when the sgRNA targeted 60-90 bases upstream of the TSS[168].

SoxS is a member of the AraC family of transcriptional regulators[168]. While there is no close homolog within *B. subtilis* to *E. coli* SoxS, the AraC family of transcriptional regulators are present. Furthermore, the function of many proteins within this family (RmgR, YbfI, YbfP, YdeC, YdeE, YesN, YfiF, YisR and YobQ) have as yet not been elucidated[17]. As such, these could be of interest for future development of CRISPRa tools for utilisation on *B. subtilis*, based on the method employed by Dong *et al*.. Given that the recognition site of SoxS upon the C-terminal domain of the RNAP α subunit is highly conserved, this system was thought to potentially be directly transferrable into *B. subtilis* or other bacterial hosts[168,169]. Unfortunately, very recently, the published PhD thesis of Dong revealed that the developed was not directly transferable to *B. subtilis*, while it was functional within *Pseudomonas putida*[78]. While Dong hypothesised that the cause of this was *B. subtilis* being intolerant to exogenous regulatory elements, the system was only tested when the GFPmut1 reporter protein was controlled by the very weak synthetic J23101 promoter. We have shown in this study that J23101 was not suitable for sgRNA expression (section 3.3.1.1), as such it would be of interest to attempt CRISPRa using this system in combination with a medium strength promoter, such as $P_{rrnB\ P1}$ (Figure 5.3).

Very recently, Liu *et al*. developed a technique by which, in *E. coli* and *Klebsiella oxytoca*, σ^{54} promoters could be activated[170]. These promoters, normally inactivated through interactions with a σ^{54} factor, become functional when a σ^{54}-activator is present. The new tool developed by Liu *et al*. localises a σ^{54}-activator to the -12 and -24 promoter elements through interactions with an sgRNA scaffold. However, this was only functional when an integration host factor (IHF)-dependent DNA loop structure is present to introduce a kink into the DNA structure, upstream of the target promoter. In *B. subtilis*, σ^{54}, encoded by *sigL,* regulates the transcription of 23 genes for adaption to cold and the utilisation of arginine, acetoin and fructose[17]. While these promoters could be attractive targets to allow dynamic control in metabolic engineering, IHF is only found in Gram negative bacteria. A possible alternative to IHF for use in *B. subtilis* could be the histone-like protein HBsu, encoded by *hbs*, which is known to be involved in DNA packaging and DNA bending and has 39% and 37% identity to *E. coli* IHF α and β subunits respectively[171].

5.4.3 The potential commercial applications of the transcriptional downregulation tool, dMAD7

Here, CRISPRi was achieved with dCas9, as well as multiplexing CRISPRi of the *amyE* and *gfpmut3* reporter genes. This represents CRISPRi on both endogenous and heterologous gene targets and as such indicates a potentially powerful tool for the identification of genes to be targeted for metabolic engineering of *B. subtilis* strains for industrial applications. When two gRNAs are combined to target either *gfpmut3* or *amyE,* the measured downregulation is not cumulative (Figure 5.20B). As the BAC0378 multiplexing results (Figure 5.20C) indicate no significant competition between the gRNAs for dMAD7, it is thought that there is a potential steric hindrance between the protospacer-dMAD7 complexes used to simultaneously target each reporter. As such, future analysis of dMAD7 enabled CRISPRi multiplexing should ensure sufficient distance between the PAM sites targeted.

Ownership over the IP rights to Cas9 and its derivatives has been intense and is becoming increasingly complex. The alternative CRISPR nuclease, MAD7 was shown in section 3.3.2 to be functional within *B. subtilis* for counterselection of cells where donor DNA had not undergone homologous recombination with the

chromosome. In this chapter we have identified the first example of a catalytically inactive variant of MAD7, dMAD7. The downregulation of *amyE* and *gfpmut3* was not as stringent as the dCas9 control included within the assay. While this is a limited comparison between dCas9 and dMAD7 given that only a single sgRNA was used for dCas9, it is in keeping with other data where DNase dead Cpf1 (ddCpf1) nucleases were not as effective for CRISPRi in other prokaryotes[74,172]. Nevertheless, this new tool represents a significant improvement in the prospects of the commercial use of CRISPR for strain construction. We propose that dMAD7 can be used to identify genes for later deletion with MAD7 without the need for a costly licence, or read-through royalties which would be incurred through the use of Cas9 or Cpf1 based tools[70].

A limitation to the dMAD7 tool developed here is its apparent lower capacity for downregulation of transcription in comparison to dCas9. Figure 5.19 shows α-Amylase to be downregulated at a lower efficiency when targeting the TTTG PAM 21 bp downstream of the start codon. This PAM sequence, on the template strand, is the same as that targeted at 11 bp and 4 bp on the template and non-template strands respectively. Furthermore, the protospacer GC% is not dissimilar to that of the 4 bp gRNA. We therefore hypothesise the cause of this lowered efficiency is due to secondary structure within the gRNA, as has previously been reported for Cas9[173,174].

Investigations into the DNA binding and release rate has shown that Cas9 does not release its cleavage product readily, while Cpf1 nucleases release the PAM distal cleavage product but not the PAM proximal product[175]. Thus, dCas9 may bind DNA in a more stable manner than ddCpf1, provided the mutations introduced to remove the catalytic activity of Cas9 or Cpf1 nucleases do not alter their capacity for DNA binding and release. Singh *et* al. also reported that and DNA was released quicker when interrogated with a catalytically inactivated Cpf1[175]. They thus theorised that DNA cleavage stabilises PAM proximal DNA binding post cleavage and that the formation of a septum (a thin barrier formed by Y201 and K1065 in *Francisella novicida* Cpf1 (FnCpf1)[176]), post DNA cleavage only, within Cpf1 could be critical to this stabilisation[175]. Given that MAD7 is part of the Cpf1 family of nucleases, this may explain dMAD7 having a lower CRISPRi efficiency and mutation of the septum formation sites may yield insights into increasing the CRISPRi efficiency of dMAD7 and ddCpf1 tools. A pairwise alignment of FnCpf1 and MAD7

reveals that Y201 is not retained, however K1065 is (MAD7 – K1021) and thus would be of interest for future mutagenesis of this tool.

Lastly, CRISPRi efficiency was analysed while targeting sites within the coding sequences of either the AmyE or GFPmut3 reporter proteins. Future experiments attempting to increase the efficiency of CRISPRi should target dMAD7 to the promoter region of the reporter protein for increased direct competition of DNA binding with the RNAP.

5.5 Conclusions

In this chapter, we have attempted to exemplify CRISPRa tools for use within *B. subtilis*, as well as developing this system for use with dCas9 to allow CRISPRi of both individual and multiplexed targets. A novel tool for downregulation was developed through the engineering of the commercially free to use MAD7 CRISPR nuclease to create the catalytically inactive dMAD7 variant. CRISPRi was performed using dMAD7 with both single, and multiplexed gene targets. This represents a promising new tool for the commercial use of CRISPR based technologies, opening this field to smaller businesses without the resources to pay for costly licences or read through royalties.

Chapter 6 – CRISPR-enabled deaminase base editing in *Bacillus subtilis*

6.1 Introduction

CRISPR tools have developed beyond the initial applications for genome editing and transcriptional regulation, as investigated in Chapter 3 and Chapter 5, to utilise dCas9 fused to effector proteins. One such tool was CRISPR guided DNA deamination, known as base editing. In such a process, nucleobase deaminases hydrolyse the amino group from deoxycytosine (C) or deoxyadenosine (A), resulting in the formation of deoxyuridine (U) or deoxyinosine (I) respectively (Figure 6.1A). These are subsequently altered to deoxythymidine (T) and deoxyguanosine (G) respectively during DNA replication, provided the deaminated base is not removed before DNA replication can utilise it as a template[81,82]. Fusing these deaminases with CRISPR nucleases, where the capacity to induce a DSB is removed, allows accurate DNA mutagenesis without the need for donor DNA to introduce the desired mutations.

In bacteria, random C to U deamination events are identified and are corrected during the base excision repair process (Figure 6.1B). Uracil-DNA glycosylase (Ung) detects and removes U by hydrolysing the N-glycosidic bond to remove the base[81]. This results in the formation of an abasic (AP) site which is in turn recognised by AP endonucleases and AP lyases, breaking the phosphodiester bond at the 5´ and 3’ sides of the AP site respectively, and forcing the repair of the site using the G on the opposite strand as a template[81]. A to I deamination reactions on the other hand are not repaired in such a manner. As such, since I is not a DNA base, when the DNA containing the I is replicated, I is recognised as the chemically similar G within the limitations of the DNA polymerase binding pocket, and incorporated into the DNA strand[80].

Deaminases capable of inducing such DNA conversions were incorporated to CRISPR systems to allow accurate targeting within a host[80,82,177]. In such tools, the deaminase, fused to either the fully catalytically inactive dCas9, or partially inactivated nCas9, can act on the single stranded DNA exposed at the distal end of the sgRNA targeting region. Here, the DNA has been relaxed by d/nCas9, and the

sgRNA has bound to the complementary strand. The ~5 base region of the opposite strand which is not bound to an sgRNA, and not buried within the DNA binding pocket of the d/nCas9, known as the editing window, is therefore free to be modified (Figure 6.2)[80,82,177].

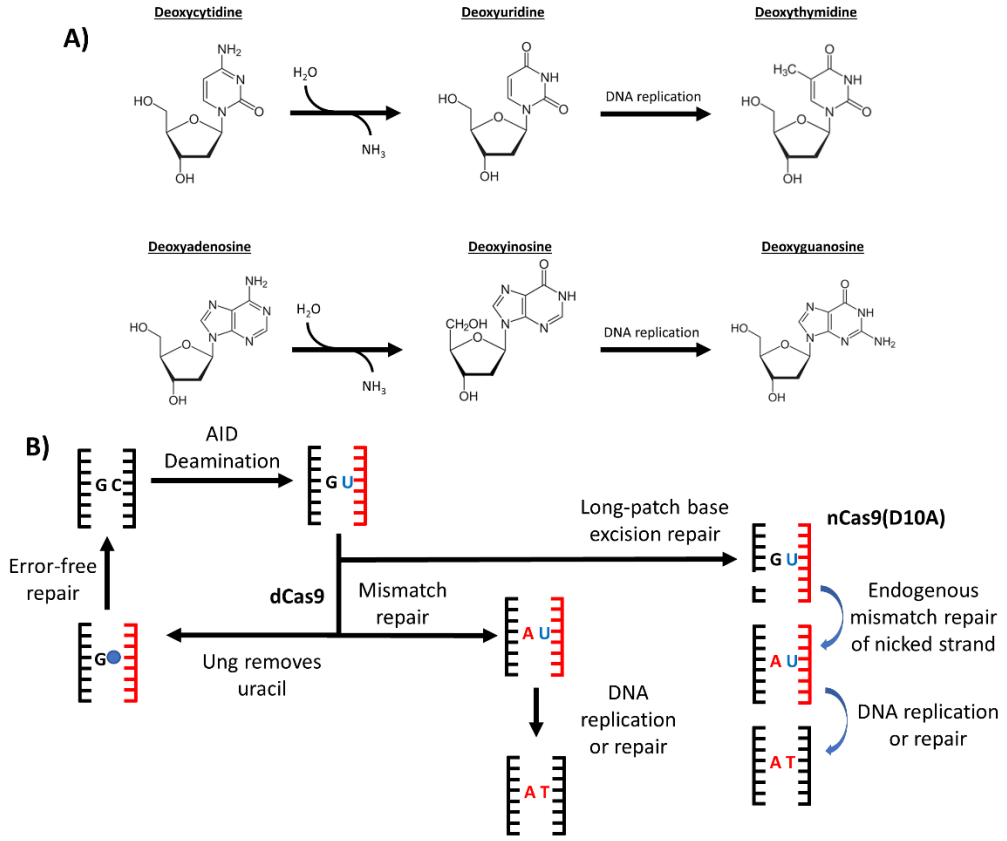


Figure 6.1 – DNA deamination and repair.

A) DNA deamination process in which a hydrolysis reaction initially creates an intermediate base, and DNA replication converts this intermediate to the final base. **B)** DNA mismatch repair process following deoxycytidine deamination by activation-induced cytidine deaminase (AID). dCas9 and nCas9(D10A) are added to indicate the route which deamination repair proceeds in the presence of these variants. Ung – uracil DNA glycosylase.

C base editing (CBE) tools were quickly developed due to cytosine deaminases known to target ssDNA having been previously identified[82]. The activation-induced cytidine deaminase (AID) enzyme from humans (encoded by *AICDA*) was identified to be an efficient C deaminase tool within human embryonic kidney (HEK) 293T cells for the alteration to T following DNA replication[79], and a sea

lamprey homolog (PmCDA1) was later taken up by Wang *et al*., for CBE within the Gram positive *Corynebacterium glutamicum*. Within *C. glutamicum*, dCas9-AID was found to have an editing efficiency of 11.2%. Utilisation of nCas9(H840A)-AID, introducing a nick on the strand targeted for deamination, did not increase the CBE efficiency. In contrast however, nCas9(D10A)-AID, cleaving the non-edited strand was found to allow 100% CBE efficiency[54]. This tool was utilised to improve the production of glutamate from *C. glutamicum* by ~2.5-fold[54].

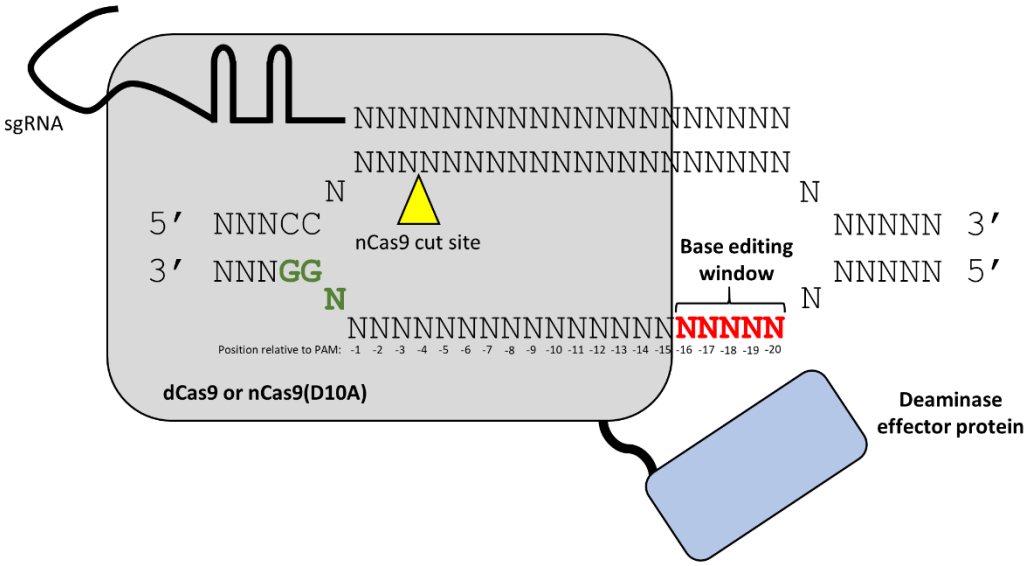


Figure 6.2 – Deaminase base editing window.

Base editing occurs 5 nt window (red) in the PAM distal region of the DNA strand which is not bound by the single-guide DNA (sgRNA). Deaminases are localised to their target by fusing them with catalytically inactive Cas9 (dCas9) or partially inactive, nickase Cas9 (nCas9). The site and strand specific DNA cleavage site for nCas9(D10A) is indicated. The 5´-NGG-3' PAM site recognised by *S. pyogenes* Cas9 is indicated (green). The positions within the protospacer region are also indicated, relative to the PAM site.

A base editing (ABE) tools were not as readily available since there were no known DNA adenosine deaminases. *E. coli* TadA is an adenine deaminase capable of converting adenine to inosine in single stranded tRNA. Gaudelli *et al*. modified the *E. coli* TadA to convert DNA A to G, via I, through seven rounds of protein evolution resulting in the construction of TadA 7.10 (referred to hereafter as TadA)[80]. This new

tool, fused to the N-terminus of dCas9, was found to be active in both *E. coli* and human cells[80]. Recently, TadA enabled base editing has also been adapted for use within *C. glutamicum,* alongside an expansion of the targeting scope through the use of the alternative nucleases xCas9 3.7 and Cas9-NG which recognise NNG/GAA/GAT/CAA and NG PAM sites respectively[84]. These represent a significant expansion from the previously exemplified *S. pyogenes* Cas9 recognising NGG PAM sites[43].

In this chapter, the first example of CBE and ABE within *B. subtilis* are reported. Differences in editing capacity and efficiency is investigated between d/nCas9 when sea lamprey AID is fused to the C-terminal, and when TadA is fused to the N-terminus. dCas9-AID was found to allow CBE, with the highest efficiency found at and either side of the -18 nt position upstream of the targeted PAM site. TadA-nCas9 was found to allow highly efficient editing with 100% of analysed colonies having at least one target base modified.

6.2 Materials and methods

6.2.1 Strains and plasmids

The strains and plasmids used in this chapter are outlined below in Table 6.1. The oligonucleotides used in this chapter can be found in Appendix C.

Table 6.1 – Strains and plasmids used in this chapter.

Strain/Plasmid	Description/Genotype	Reference
Strains		
B. subtilis 168	*trpC2*	Laboratory stock
BAC0353	*B. subtilis* 168 with pBAC0184	This chapter.
BAC0387	*B. subtilis* 168 with pBAC0223	This chapter.
BAC0388	*B. subtilis* 168 with pBAC0169	This chapter.
BAC0389	*B. subtilis* 168 with pBAC0170	This chapter.
BAC0390	*B. subtilis* 168 with pBAC0171	This chapter.

BAC0391	*B. subtilis* 168 with pBAC0172	This chapter.
BAC0397	*B. subtilis* 168 with pBAC0192	This chapter.
Plasmids		
pdCas9-AID	Tc^R; p15a ori; *cat*; P_{trc}; *dCas9-AID*	Laboratory stock
pnCas9(D10A)-AID	Tc^R; p15a ori; *cat*; P_{trc}; *nCas9(D10A)-AID*	Laboratory stock
pTFG020A	pBR322 ori; *bla*; P_{CMV}; *tadA-dCas9*	Laboratory stock
pCMV-ABE7.10	pBR322 ori; *bla*; P_{CMV}; *tadA-nCas9(D10A)*	[80]
pHT01	*E. coli/B. subtilis* shuttle vector carrying P_{grac} and *lacI, bla*, *cat*	MoBiTec
pBAC0001	pHT01 with SapI sites removed	Section 3.2.1.1
pBAC0008	pBAC0001 with rrnB T1 T2; sgRNA Cas9 handle; protospacer cloning site; P_{veg}	Section 3.2.1.1
pBAC0015	pBAC0008 with *cas9* (*S. pyogenes*). Cas9 expression regulated by the P_{grac}	Section 3.2.1.1
pBAC0018	pBAC0008 with d*cas9* (*S. pyogenes*). dCas9 expression regulated by the P_{grac}	Section 5.2.1.1
pBAC0041	pBAC0015 with sgRNA targeting 25 bp downstream of the start codon within *B. subtilis* 168 *amyE*	Section 3.2.1.1
pBAC0169	pBAC0041 with *tadA-dcas9* replacing *cas9*	This chapter.
pBAC0170	pBAC0041 with *tadA-ncas9* replacing *cas9*	This chapter.
pBAC0171	pBAC0041 with *dcas9-AID* replacing *cas9*	This chapter.
pBAC0172	pBAC0041 with *ncas9-AID* replacing *cas9*	This chapter.
pBAC0175	pBAC0015 with *dcas9-AID* replacing *cas9*	This chapter.
pBAC0176	pBAC0015 with *ncas9-AID* replacing *cas9*	This chapter.
pBAC0184	pBAC0041 with *dcas9* replacing *cas9*	Section 5.2.1.1

pBAC0192	pBAC0175 with sgRNA targeting 40 bp downstream of the start codon within *B. subtilis* 168 *nicK*. oMAP0682/0683	This chapter.
pBAC0193	pBAC0176 with sgRNA targeting 40 bp downstream of the start codon within *B. subtilis* 168 nicK. oMAP0682/0683	This chapter.
pBAC0223	pBAC0018 with sgRNA targeting 40 bp downstream of the start codon within *B. subtilis* 168 *nicK*. oMAP0682/0683	This chapter.
pBAC0224	pBAC0041 with *ncas9(D10A)* replacing *cas9*	This chapter.

6.2.1.1 Plasmid construction

The construction of plasmid pBAC0008 using the inABLE® assembly technique is described in detailed in Section 3.2.1.1. The introduction of *cas9* into pBAC0008 (yielding pBAC0015) is also described in detailed in Section 3.2.1.1. The introduction of *dcas9* into pBAC0008 and pBAC0041 (yielding pBAC0018 and pBAC0184 respectively) is described in detailed in Section 5.2.1.1.

pBAC0169/0170/0171/0172/0224 were prepared by digesting pBAC0041 with BamHI-HF and XbaI to remove *cas9* and subsequently ligating with BsaI and XbaI digested PCR products prepared using the following respective templates and PCR primers: *tadA-dcas9* from pTFG020A with oMAP0638/0639; *tadA-ncas9(D10A)* from pCMV-ABE7.10 with oMAP0638/0639; *dcas9-AID* from pdCas9-AID with oMAP0640/0641; *ncas9(D10A)-AID* from pnCas9(D10A)-AID with oMAP0640/0641; *ncas9(D10A)* from pBAC0176 with oMAP0073/0074. The same digested *TadA-ncas9(D10A), dcas9-AID* and *ncas9(D10A)-AID* PCR products were ligated to the BamHI-HF and XbaI digested pBAC0015 backbone to yield plasmids pBAC0175/0176 respectively. pCMV-ABE7.10 was a gift from David Liu (Addgene plasmid #102919)[80]. pTFG020A was a gift from Changhao Bi in which *tadA-nCas9(D10A)* from pCMV-ABE7.10 had been modified to *tadA-dCas9* and inserted into a *E. coli* expression vector also containing an sgRNA expression cassette. pdCas9-AID and pnCas9(D10A) were also gifts from Changhao Bi in which the relevant fusion gene was ligated with the low copy *E. coli* plasmid, pACYC184 (NEB, USA).

Plasmids used for analysis of the base editing fusion proteins dCas9-AID or nCas9-AID were prepared from the pBAC0175 or pBAC0176 parental plasmids respectively, using phosphorylated and annealed oligonucleotide pairs as described in section 2.5.4.1. Control plasmids were prepared from pBAC0018 in a similar manner. Information on the oligonucleotides used in construction, target gene and PAM site for the following CRISPR plasmids used in this chapter can be found in Table 2.12: pBAC0192; pBAC0193, pBAC0223.

6.2.1.2 Strain construction

All strains described in Table 6.1, other than *B. subtilis* 168, were prepared by making *B. subtilis* 168 naturally competent (section 2.4.2) and transforming with the plasmid also described in Table 6.1.

6.3 Results

6.3.1 Deaminase base editing as a method for DNA modification without the need for a double strand break within *Bacillus subtilis*

To exemplify deaminase base editing within *B. subtilis*, the fusion genes *tadA-dcas9*, *tadA-ncas9*, *dcas9-AID* and *ncas9-AID* (provided by Changhao Bi – Tianjin, China) were used to replace *cas9* in pBAC0041, creating plasmids pBAC0169/0170/0171/0172 respectively. Plasmid pBAC0041 had previously been found to efficiently allow the targeting of the CRISPR-Cas9 system to *amyE* for counterselection by DNA cleavage within cells which had not undergone homologous recombination of the linear dDNA fragment conveying mutations to *amyE* alongside a synonymous PAM site mutation (section 3.3.1.4). pBAC0041 was also modified, replacing *cas9* with *dcas9,* to allow the down regulation of *amyE* mRNA levels for a reduced level of α-Amylase activity (section 5.3.3.3). Assuming the deaminase base editing window (Figure 6.2) is the same in *B. subtilis*, it was hypothesised that during treatment with TadA, two A bases, -16 and -20 nt from the PAM site, would be available for deaminase base editing to G, while with AID treatment one C, -18 nt from the PAM site, would be available for conversion to T (Figure 6.3).

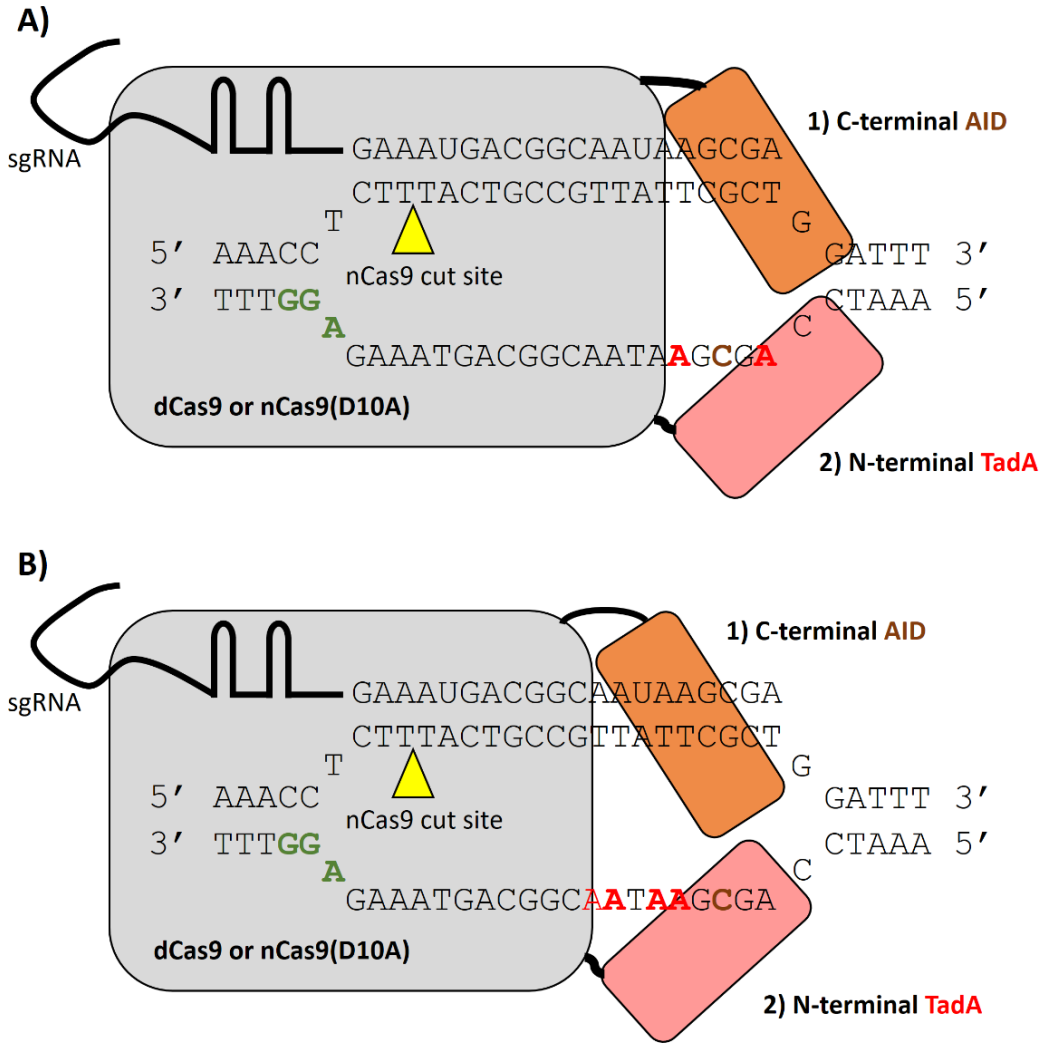


Figure 6.3 – Deaminase base editing window identification in pBAC0041 based plasmids.

A) The proposed 'window' in which deaminase base editing can occur. **B)** The observed 'window' in which deaminase base editing can occur. Deaminase proteins are fused to dCas9 or nCas9. **1)** Activation-induced cytidine deaminase (AID) was C-terminally fused or, **2)** the modified E. coli TadA, ABE7.10[80], was N-terminally fused to enable the targeted deamination reaction. The nCas9(D10A) strand specific cut site is indicated with the yellow triangle. Proposed or observed bases susceptible to deamination by TadA (red) or AID (brown) are indicated. Bold bases indicate where base editing was common, while non-bold indicate where base editing was irregularly observed. The PAM site targeted is indicated in green.

6.3.1.1 AID enabled CBE within *B. subtilis*

Upon transformation of *B. subtilis* 168 with pBAC0171 and pBAC0172, carrying dCas9-AID and nCas9-AID respectively, transformants were only obtained with dCas9-AID. Transformation of the control plasmids pBAC0184 and pBAC0224, expressing dCas9 or nCas9 respectively, and targeting the same PAM site as pBAC0171 also showed nCas9 to be lethal to the cell in the absence of AID while dCas9 yielded similarly high CFUs as dCas9-AID – a similar result to that observed in *E. coli* by Banno *et al*.[178], but contrasting the results obtained by Wang *et al*. where nCas9-AID was used to increase base editing efficiencies in the Gram positive *C. glutamicum*[54].

Colony PCR (cPCR) and sequencing was used to analyse three transformants of *B. subtilis* with dCas9-AID targeting *amyE* (pBAC0171, strain BAC0390). A mixed signal was observed in all the sequencing chromatograms at the targeted base within the base editing window. This indicated the colony was made up of cells with a mixed genotype, and deaminase base editing of C to T with AID was achievable. No other C within the sequencing reactions was found to be edited, indicating the deaminase reaction is targeted with the CRISPR-Cas9 system and not random.

To determine the proportion of cells edited within these initial colonies, two were streaked out to isolate individual cells from the transformant. cPCR and sequencing was subsequently used to analyse eight CFUs of the sub-population in the same manner. No homogenous non-edited (WT) colonies were observed. 55% of the colonies exhibited a mixed C/T genotype, while 45% were found to only have the signal indicating T (Figure 6.4). The lack of a statistically significant difference between the edited and WT colonies in the transformants indicates that AID base editing is likely to occur during DNA replication with one daughter cell being edited and the other non-edited. This hypothesis also implies the lack of editing prior to colony formation commencing on the agar plate as editing during the recovery phase would enable the full colony being of the single, edited genotype.

To further characterise the editing window of this new *B. subtilis* CBE tool, pBAC0192 (dCas9-AID) and pBAC0193 (nCas9-AID) were constructed to target a region of seven consecutive C bases (Figure 6.5A) within the *nicK* gene, encoding a non-essential DNA relaxase[17]. Once again, no transformants were obtained with the nCas9-AID plasmid. Two dCas9-AID transformants (strain BAC0397) underwent

cPCR and sequencing, revealing a mixed genotype at positions -17 and -19 from the PAM site, while position -18 appeared to be almost fully edited in both colonies, with only a very weak C signal observed in one of the colonies (Figure 6.5B).

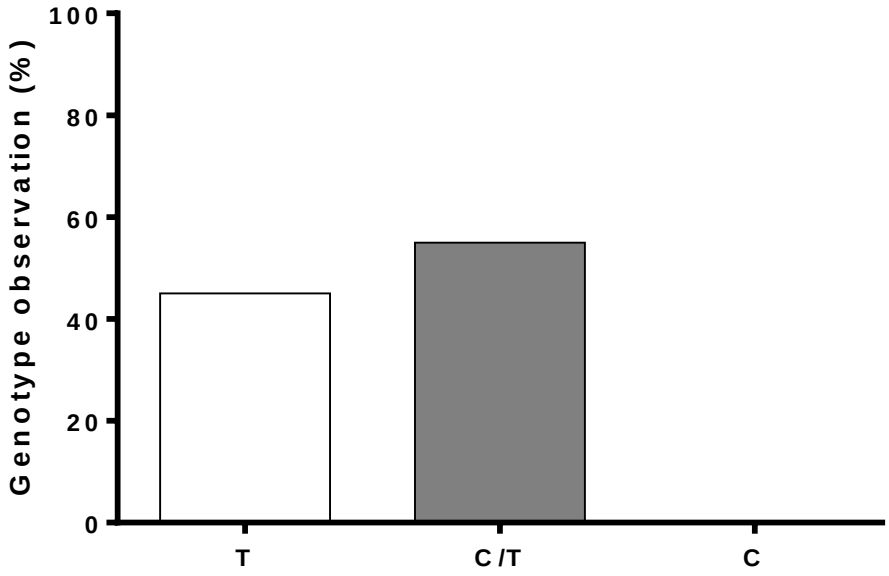


Figure 6.4 – AID enabled C to T deaminase base editing efficiency within *B. subtilis* transformants.

Bars represent the genotype observed (%) following sequencing of the targeted region of eight subpopulations from original transformants targeting dCas9-AID to *amyE* in strain BAC0390. Levels of fully edited (T), mixed (C/T) and WT (C) traces are indicated.

In an attempt to fully edit the additional target bases, each colony was grown in LB supplemented with chloramphenicol and 1 mmol L^{-1} IPTG to mid-log phase and subsequently grown for 24 hours in fresh media of the same composition. IPTG was included for dCas9-AID expression. Following 24 hours of growth, colonies were isolated by spreading on LB agar plates supplemented with chloramphenicol only. Eight subpopulation colonies were analysed from each culture by cPCR and sequencing. All chromatographs for one culture set were found to be of single genotype reads, while the other set exhibited one mixed genotype, at position -17 only, and two failed sequencing reactions. Using the 8 and 5, single genotype chromatograms, a base editing efficiency was established for positions -19 (16.3%), -18 (100%), -17 (90%) and -16 (6.3%) as shown in Figure 6.5C. cPCR and

sequencing of three colonies for each of three transformants expressing dCas9 only and grown in the same manner for 24 hours showed no modifications at either the *amyE* (strain BAC0353) or *nicK* (strain BAC0387) target sites. Thus, CBE was confirmed as having been caused by the AID fused to dCas9.

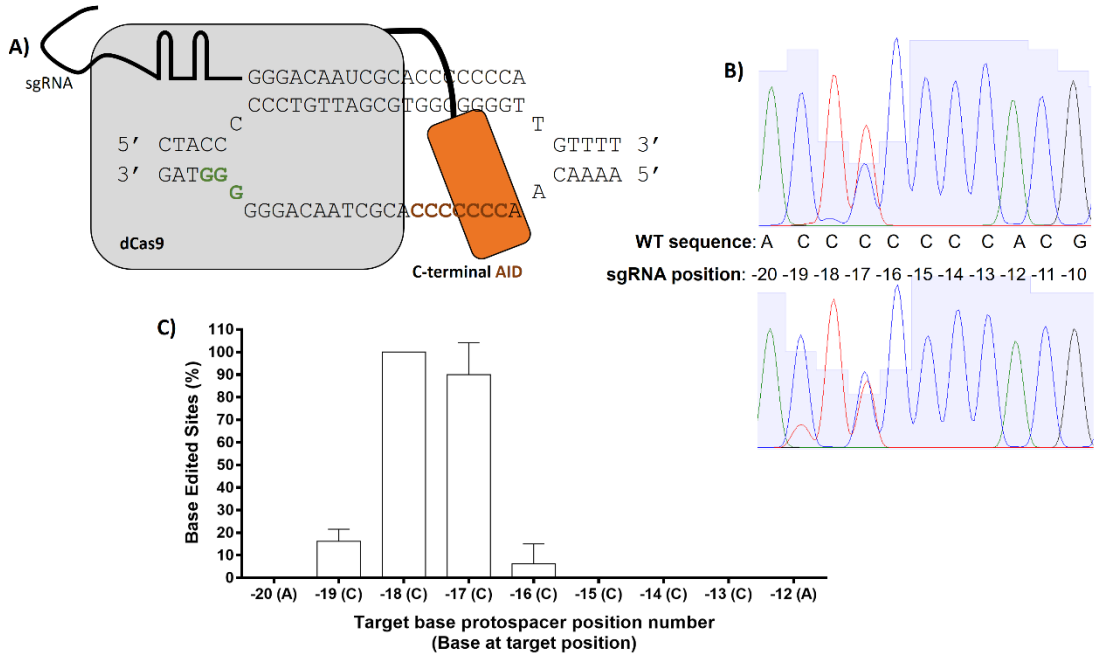


Figure 6.5 – dCas9-AID deaminase base editing 'window' analysis.

A) Targeting schematic of seven consecutive C bases (brown) by dCas9-AID to analyse the deaminase base editing window. The PAM site targeted is indicated in green. **B)** Sequencing chromatogram following analysis of two initial transformants for dCas9-AID targeting *nicK* in strain BAC0397 (C – blue; G – black; A – green; T – red). The starting wild-type (WT) DNA sequence and its position within the sgRNA are indicated. **C)** Bars represent dCas9-AID base editing efficiencies at each targeted base within region covered by the sgRNA. Error bars indicate the standard deviation between subpopulation genotypes of the initial starting colonies.

6.3.1.2 N-terminally fused TadA enabled ABE within *B. subtilis* and masks the lethality of nCas9

The readily available fusion constructs, TadA-dCas9 and TadA-nCas9 were used to replace Cas9 in pBAC0041, yielding plasmids pBAC0169 and pBAC0170 respectively. Here, the evolved *E. coli* TadA deaminase was fused to the N-terminus

of dCas9 or nCas9(D10A). The resulting plasmids, also expressing a sgRNA specific for *amyE*, both enabled the formation of colonies following transformation of *B. subtilis* 168. While the C-terminal fusion of dCas9-AID had enabled the formation of colonies, nCas9-AID had not. Additionally, nCas9 alone had also conveyed a similar lethality when targeting *amyE*. Therefore, the fusion of TadA to the N-terminus of nCas9 was hypothesised to convey a masking effect to the lethality of the catalytic activity for the HNH domain, retained within nCas9(D10A).

To determine if TadA enables deaminase base editing within *B. subtilis*, three transformants were first screened by cPCR and sequencing of the *amyE* region targeted by TadA-dCas9 and TadA-nCas9 (Figure 6.3) in strains BAC0388 and BAC0389 respectively. Only the WT sequence was identified in all chromatograms. Each colony was grown in LB supplemented with chloramphenicol and 1 mmol L^{-1} IPTG to mid-log phase and subsequently grown in fresh media of the same composition. Single colonies were isolated on LB agar supplemented with chloramphenicol only following 16, 24 and 48 hours of growth. Six of these colonies for each original transformant were screened for deaminase base editing at each timepoint by cPCR and sequencing.

The results indicate that TadA is functional for deaminase base editing within *B. subtilis* 168 when fused to the N-terminus of either dCas9 or nCas9 (Figure 6.6). A maximum average editing rate of 33% was observed between the biological replicates, after 48 hours of induction, at position -15, when treated with TadA-dCas9. In contrast, the same position had an average editing rate of 83% when treated with TadA-nCas9. This, along with position -16 after 24 hours, was also the maximum base editing efficiency achieved at a given protospacer position with TadA-nCas9. A similar increase in efficiency between the two fusion proteins was noted at positions -13 and -16 for all timepoints. A clear relationship was therefore observed between an increase in the observed base edited sites, and the use of TadA-nCas9 when compared to TadA-dCas9. This indicates a level of selection conveyed by nCas9(D10A) and thus we determined that TadA does not fully mask the catalytic activity of the HNH domain.

Position -12 appears to be the inner extremity of the deaminase base editing window under the tested conditions with a 66% drop in efficiency at 24 hours between positions -13 and -12 when treated with TadA-nCas9, and no editing observed at all for position -12 following TadA-dCas9 treatment. The outer extremity

of the editing window was not identified in a similarly robust manner since no drop off in efficiency was observed. However, since another A was available at position -20 within the protospacer, and no editing was observed here in any sample with either fusion protein, this appears to be the minimum cut off point for the editing window of TadA-d/nCas9.

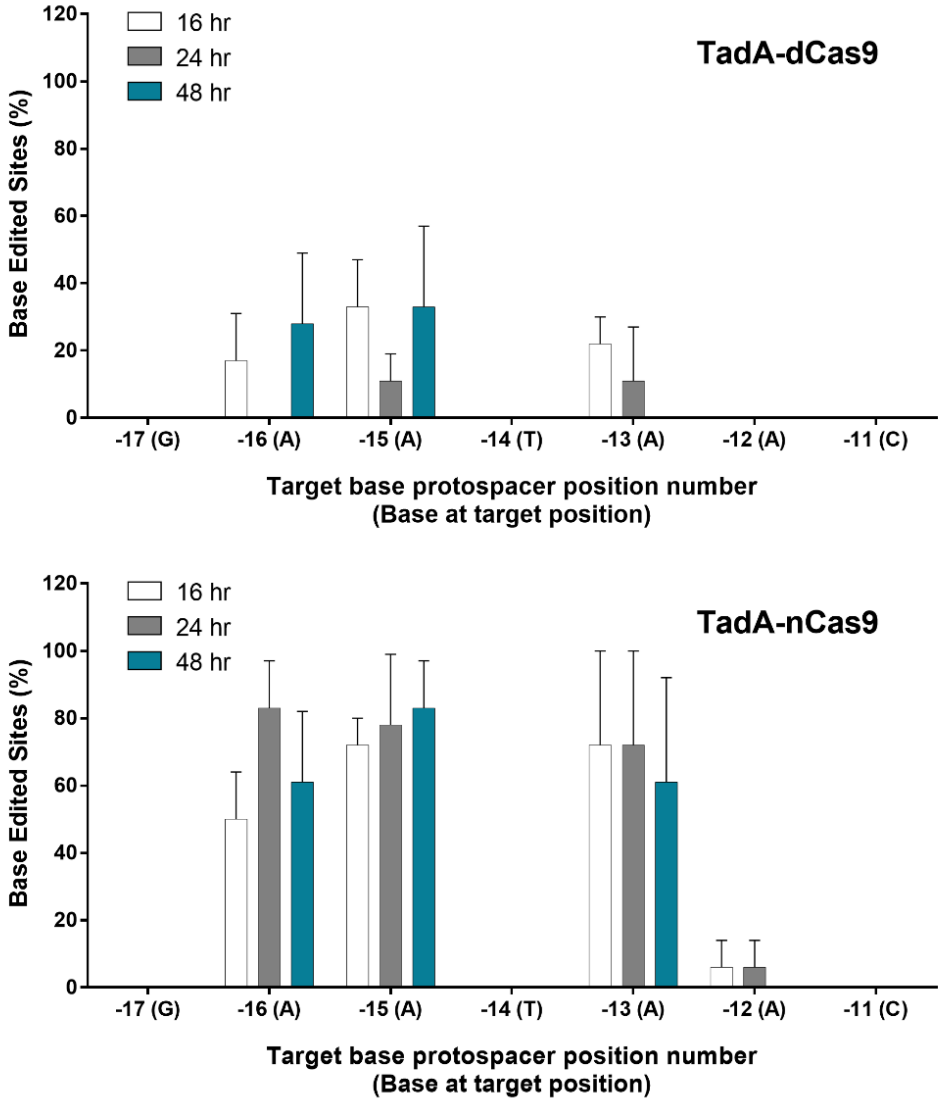


Figure 6.6 – TadA-dCas9 and TadA-nCas9 base editing levels at the *amyE* locus.

Base editing levels were determined by colony PCR and sequencing of isolated colonies following 16 (white), 24 (grey) and 48 (blue) hours of growth. Bars represent the rate (%) where base editing has occurred and is indicated on the sequencing chromatogram across the analysed colonies. Error bars indicate the standard deviation between subpopulation genotypes of the initial starting colonies.

While TadA-dCas9 was found to successfully enable deaminase base editing, it was striking that of the 18 sequencing reactions (three biological replicates with analysis of six individual colonies for each) following 16, 24 and 48 hours of induction, only 7, 11 and 9 chromatograms respectively did not reveal any mixed genotypes by the presence of a mixed signal. TadA-nCas9 on the other hand yielded 17/18 homogenous chromatograms at each sampling point. Analysis of the single genotype chromatograms only for each sample point further emphasised the benefit of fusing TadA to nCas9 since no WT colonies were identified, whereas non-edited colonies made up 70% of those treated with TadA-dCas9 (Figure 6.7). Furthermore, there appears to be a preference for two base editing events when treated with TadA-nCas9, with 73% showing this genotype when all 16, 24 and 48 hour sample sets are combined. Of these, positions -16, -15 and -13 were edited in 59%, 78% and 62% of colonies respectively. Position -12 was not edited in any case where only two sites were modified. As such, there appears to be a slight preference for the editing of position -15 by TadA-nCas9.

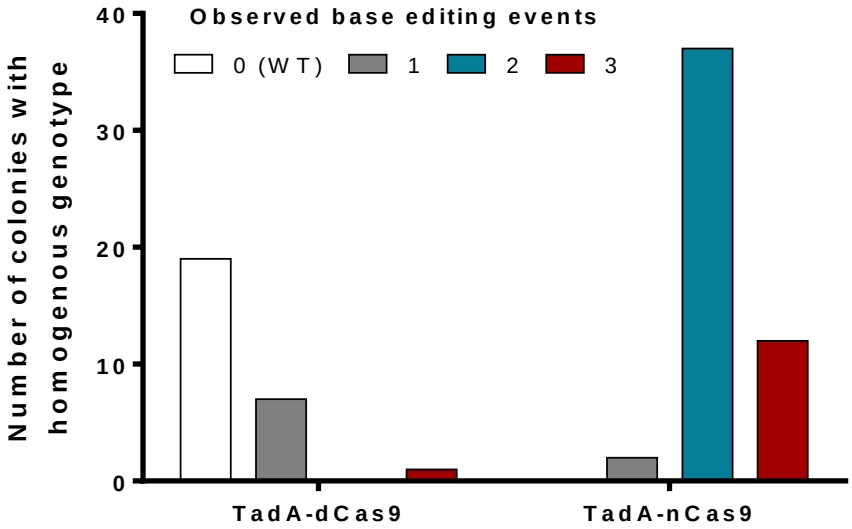


Figure 6.7 – Number of TadA deaminase base editing events observed in chromatograms exhibiting a single genotype.

Bars represent the total number of single genotype chromatograms obtained following colony PCR and sequencing of all colonies screened following 16, 24 and 48 hours of growth. Bars are split between chromatograms where no base editing (WT) was observed (white) and where 1 (grey), 2 (blue), or 3 (red) base editing events were observed.

It was hypothesised that TadA only partially masks the catalytic activity of the Cas9 HNH domain when fused to the N-terminus of nCas9(D10A) following the clear increase in base editing efficiency in comparison to TadA-dCas9 (Figure 6.6). Due to nCas9-AID and nCas9 on its own being lethal to *B. subtilis* 168, the growth of BAC0388 and BAC0389 (*B. subtilis* 168 maintaining pBAC0168 or pBAC0170, as well as expressing TadA-dCas9 and TadA-nCas9 respectively) was investigated to determine if the presence of TadA-nCas9 conveyed a negative effect on growth rate. *B. subtilis* 168 was re-transformed with pBAC0169 and pBAC0170 and transformants spread on LB agar supplemented with chloramphenicol only. Three transformants for each strain were grown to mid-log phase in LB supplemented with chloramphenicol only, before the growth for each was continuously monitored by light backscatter, within a 50 mL shake flask culture, using the cell growth quantification system (Aquila Biolabs – section 2.2.5). The same medium was utilised with the further supplementation of 1 mmol L^{-1} IPTG for TadA-dCas9 and TadA-nCas9 expression induction.

The results show a slower initial growth rate for the cultures in which TadA-nCas9 was expressed. This subsequently increases to a maximum of 0.29 A.U./min following 6.38 hours of growth, which is comparable to the 0.27 A.U./min growth rate at 3.13 hours when TadA-dCas9 is expressed (Figure 6.8). We hypothesise this reduced initial growth rate to be caused by nicking of DNA by nCas9 at reduced efficiency, and therefore not lethal to the whole population (as with nCas9-AID or nCas9 only), due to the presence of TadA.

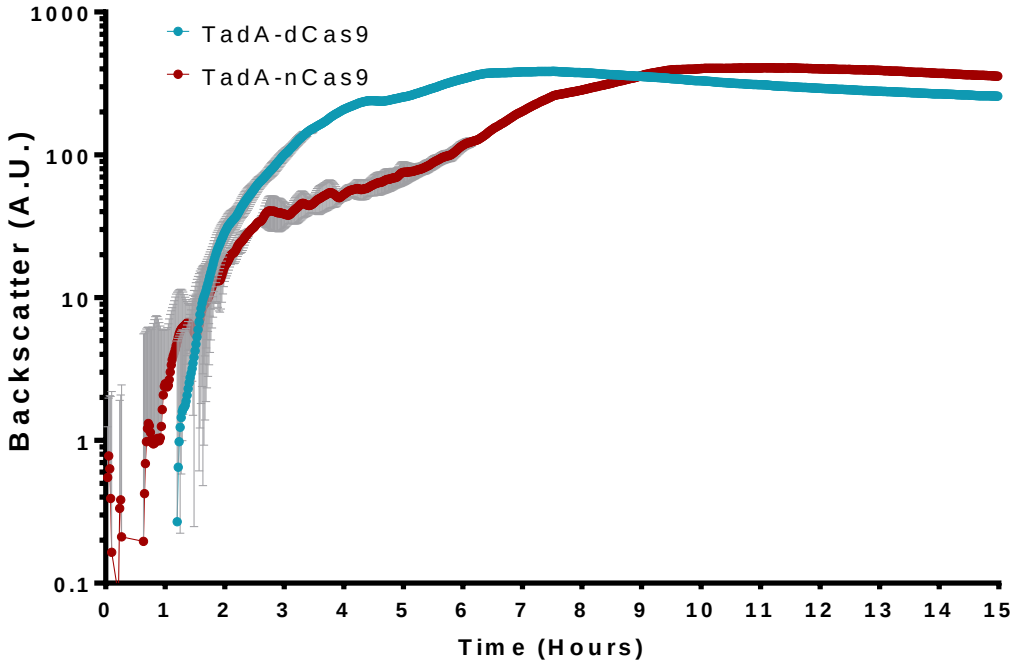


Figure 6.8 – Growth curves of strains expressing TadA-dCas9 or TadA-nCas9.

The growth of strains BAC0388 (blue) and BAC0389 (red) containing pBAC0169 (TadA-dCas9) and pBAC0170 (TadA-nCas9) respectively were monitored with light backscatter using the Cell Growth Quantifier (Aquila Biolabs – section 2.2.5) over 15 hours with a measurement taken every 60 seconds. Grey error bars represent the standard deviation between 3 biological replicates.

6.4 Discussion

6.4.1 dCas9-AID enable CBE activity is centred around nucleotide 18 within the protospacer

While the targeting of *amyE* with dCas9-AID only had a single target nucleotide, it appears serendipitous that this target was at nucleotide -18 within the protospacer region. Analysis of an alternative site within *nicK*, containing 7 consecutive target C nucleotides, indicated a significantly reduced activity at positions -16 and -19, with highly efficient editing observed at both positions -17 and -18. This is in contrast to results shown by Wang *et al*., where editing frequencies of ~40-70% were observed across positions -16 to -20[54].

Notably, here we have only tested the C-terminally fused dCas9-AID for CBE. It would be of interest to determine if N-terminally fused AID-dCas9 allowed a

similar level of CBE efficiency. Furthermore, investigation of an N-terminally fused AID-nCas9 would be of interest to determine if it enabled a reduction in the efficiency of nCas9 DNA cleavage, removing the absolute lethality of nCas9 in a similar manner to TadA-nCas9. This could allow CBE within *B. subtilis* at a greater editing efficiency, a higher proportion of isolated colonies with a homogenous genotype, and the isolation of colonies where bases are modified which are not currently with the dCas9-AID tool developed here.

6.4.2 TadA-nCas9 enriches the selection of cells which have undergone ABE

We have observed that nCas9 is lethal to *B. subtilis* 168 when targeted to the chromosome both on its own and also when AID is fused to its C-terminus (nCas9-AID). This has recently been confirmed and exploited by Liu *et al*. to establish a multiplexed editing technique within *B. subtilis* 168[116]. Interestingly, we also observed that when the ABE deaminase, TadA, was fused to the N-terminus of nCas9, the absolute lethality of nCas9 was lost. The partial toxicity which remained (Figure 6.8) was harnessed to allow the selection and enrichment of cells which had undergone successful base editing.

Peak growth rate for the TadA-dCas9 expressing strain, BAC0388, was observed at 3.13 hours (Figure 6.8), while that of BAC0389 was observed at 6.38 hours (0.27 and 0.29 A.U./min respectively) after recovering from an initial comparatively poor growth rate. Given that transformants had shown no editing to occur prior to growth in the presence of IPTG, we propose the recovery in growth rate was due to the successful editing of the target sites within the BAC0389 population. This editing in turn prevents the full hybridisation of the sgRNA and subsequently stops the final conformational changes within nCas9, required for the HNH domain to cleave its DNA strand (Figure 1.1)[179]. Furthermore, only two colonies were identified as having only a single base edited (Figure 6.7), both of which were identified following 16 hours of growth, further indicating an enrichment over time of cells containing a modified target region with ≥2 bases edited.

It is interesting that it appears that the *amyE* targeting sgRNA screened here causes a loss in the ability for nCas9 to cleave the DNA, when ≥2 sites within the PAM-distal region of the target DNA are modified. It is known that modification of

sites within the seed region (8 nt adjacent to the PAM site within the protospacer) significantly or totally inhibits the effectiveness of the sgRNA to bind the target locus[179]. Changes to the PAM-distal end on the other hand are known to reduce activity, with up to 5 changes being tolerated within *E. coli*[180] before full loss of cleavage. It is interesting therefore that such an apparent drop in cleavage capacity (observed through the re-establishment of a comparable growth rate in Figure 6.8) is caused by a modification at 1-3 target nucleotides in *B. subtilis* 168, and could indicate the sgRNA utilised is not optimally designed.

6.4.3 TadA-nCas9 enables a greater editing efficiency in *B. subtilis* than in *C. glutamicum*

Wang *et al*. found that the editing window for TadA-nCas9(D10A) within the Gram-positive *C. glutamicum* was from positions -13 to -17 within the protospacer region of 14 targets[84]. However, the editing efficiency observed by Wang *et al*. only ranged from 2.5% to 44.2%. This tool therefore, albeit only at a single target, confers a greater editing capacity within *B. subtilis* with an editing range of 6% to 83% across positions -12 to -16. Wang *et al*. speculated that A to I editing events are repaired differently in prokaryotes than eukaryotes given the drop in editing efficiency in comparison to humans (~50%)[80], and that this was the cause of lower editing efficiencies. The results presented here do not support this theory for prokaryotes in general, and may even be isolated to *C. glutamicum* and other close relatives, since *B. subtilis* is the Gram-positive model organism. However, more sgRNAs must be examined within *B. subtilis* and other prokaryotes to draw robust conclusions on this matter.

6.5 Conclusion

In this chapter we have established the first example of deaminase base editing within *B. subtilis.* Both CBE and ABE systems were exemplified with initial experiments to characterise the editing window for both systems carried out. TadA fused to the N-terminus domain of nCas9 masked the lethality observed with nCas9 and nCas-AID, yielding a highly efficient tool for A-G modifications in *B. subtilis*.

Chapter 7 – Conclusions and future work

In this study we have developed a flexible toolbox enabling the use of CRISPR based genome editing techniques within the Gram-positive model organism, *Bacillus subtilis*. Our system consists of a single plasmid for IPTG-inducible expression of a nuclease, and constitutive expression of a gRNA. This plasmid is co-transformed alongside a dDNA which is PCR amplified or constructed by OE-PCR. Transcriptional regulation was achieved using a catalytically inactivated variant of the CRISPR nuclease, expressed from a similar single plasmid system, and used to transform *B. subtilis* in the absence of dDNA. Furthermore, we developed a tool, the use of which had previously not been exemplified in *B. subtilis*, for genome editing via DNA deamination in the absence of dDNA as a template for modifications. We have focused on the industrial applications of these tools, demonstrating protein engineering *in situ*, creating a library of knockouts to increase heterologous protein production, and developing genome editing and transcriptional interference with a commercially attractive CRISPR nuclease.

We have demonstrated that the widely used CRISPR nuclease, *S. pyogenes* Cas9, and the potentially industrially significant alternative CRISPR nuclease, MAD7, are both efficient for *B. subtilis* genome editing. Cas9 was primarily used throughout this study and successfully enabled point mutations and DNA insertions, while also enabling the knock-out of genes, and knock-in of promoters, during the construction of strains used for analysis of transcriptional upregulation. MAD7 was investigated for its genome editing capacity within *B. subtilis* following its release by Inscripta. A comparable editing rate was observed in comparison to Cas9, showing that MAD7 could, along with the required gRNA modifications, directly replace Cas9 in our editing system. MAD7 was subsequently used for the construction of a strain library intended for the production of heterologous proteins. This was constructed though frame-shift mutations of extracellular proteases. These strains could be further developed for the knock-out of additional genes such as *amyE*, *spoIIE* and *srfAC*. Such strains would have a greater applicability towards industrial fermentation processes as they would have less extracellular proteins interfering with product purification, an inability to sporulate, and a reduced liability to foam which would result in the loss of product.

We have provided evidence that homologous recombination of the dDNA prevents nuclease activity through disruption of the PAM site, or protospacer region. This is true for both Cas9 and MAD7 CRISPR nucleases. In the context of the assay used to display this, the CRISPR nuclease was not required since the curing of tryptophan auxotrophy in *B. subtilis* was used as a selection mechanism on a minimal medium. This feature enabled the clarification that homology directed repair of the DNA DSB with the dDNA was not a significant factor in CFU formation following co-transformation with a plasmid targeting the *trpC2* lesion. However, where such a selection mechanism is unavailable, the CRISPR system acts as a counterselection tool against cells which have not undergone homologous recombination of the dDNA. In the absence of the CRISPR system, transforming *B. subtilis* with only dDNA and spreading on plates without an antibiotic selection pressure, it can be expected that a similarly high proportion of the cells will be transformed. However, significant dilution of the transformation culture would be required to isolate individual colonies. Thereafter, screening of transformants would require colony PCR and sequencing of many colonies to potentially only identify a single successfully edited colony. Thus, following co-transformation with dDNA (which does not contain an antibiotic resistance gene for the selection of successful genomic integrants), the CRISPR system drastically reduces the number of transformants to be screened following counterselection with the nuclease, as well as antibiotic selection for the presence of the CRISPR plasmid.

We additionally tested the capacity for AsCpf1 to enable genome editing in *B. subtilis*. AsCpf1 in combination with a non-targeting gRNA was found to be toxic to the cell, with very few transformants obtained. A similar non-targeting gRNA was used with MAD7 and Cas9 as a control for transformation efficiency determination and toxicity to *B. subtilis*. As MAD7 is within the Cpf1 family of nucleases and was found not to be toxic to *B. subtilis* without a targeting gRNA, we hypothesise that this toxicity was not due to the AsCpf1 nuclease being intrinsically toxic to the cell. Instead, we suggest the human codon optimised gene we used inhibited growth due to pauses in the mRNA translation causing an accumulation of misfolded or incomplete insoluble protein, known to be toxic to the cell. Future experiments should utilise either the WT sequence of *AsCpf1*, or an equivalent which is codon optimised for expression within *B. subtilis*.

We used this co-transformational CRISPR approach to rapidly develop an improved variant of the subtilisin E alkaline protease with greater thermotolerance

(1.4 °C) and activity (46.5%) under neutral conditions. These were modified *in situ*, incorporating a salt-bridge known to aid thermotolerance in the *B. clausii* homolog, M-protease. This is an exemplification of our system for the modification of established production strains for any given protein of interest. dDNA libraries can be constructed through OE-PCR to introduce a mutation within the PAM site, followed by error-prone PCR to introduce random modifications within the sequence. Alternatively, OE-PCR products could be constructed using oligonucleotide pools where a specific region is targeted for a designed set of modifications, for example an enzyme binding pocket. Such modifications have the capacity to further increase the efficiency of enzymes currently being produced and make established industrial fermentation processes more profitable.

The co-transformational approach utilised here worked well in *B. subtilis* 168, however its limitations began to become apparent when BAC0288 (*B. subtilis* 168 *ΔbglS::aph(3')-lIIa*-P_{veg}-*gfpmut3*) was utilised. An ~10-fold drop in transformation efficiency was observed, severely impacting on the number of transformants obtained. This may replicate the lower transformation efficiencies which have been observed within some industrial strains that have been significantly modified, and as such indicates a limitation with our approach. In cases where strains of *B. subtilis* with a poor transformation efficiency are utilised, it may be more appropriate to utilise other systems, for example that of Altenbuchner[63], where the dDNA is contained within the editing plasmid alongside the sgRNA and Cas9 nuclease expression cassettes. However, for strains with efficient natural competency, our system represents a rapid and versatile approach for genomic modifications.

MAD7 was successfully engineered to carry three amino acid modifications (D877A, E962A and D1213A). Through survival data, we interpret this variant to be the first catalytically inactive MAD7 variant – dMAD7. Further development and characterisation, through purification and *in vitro* assays, would be of great interest to directly demonstrate that this dMAD7 variant lacks the capacity to cleave DNA. We have shown here that multiplexing is possible with dMAD7, however we have not confirmed if MAD7, or our engineered dMAD7, has the capacity to mature its own pre-crRNA in a similar manner to other Cpf1 nucleases[181]. In bacterial hosts, dMAD7 may, like Cas9, combine with RNase III to mature the pre-crRNA. This process is not the same in mammalian hosts and therefore it is desirable to know if dMAD7 can mature its own crRNA array to allow multiplexing in higher eukaryotes. If dMAD7 cannot mature its own pre-crRNA, it may be desirable to return some of

the modified amino acids to their WT form in an effort to obtain a DNase inactivated variant of MAD7 which can cleave and mature pre-crRNA. This is the case, for example, in AsCpf1 E993A which retains the capacity to cleave and mature its own pre-crRNA array, while removing its capacity for DNA cleavage[74]. However, the putative residues for pre-crRNA processing in AsCpf1 are His800, Lys809 and Lys860 (corresponding to His759, Arg768 and Lys830 within MAD7). The corresponding residues in *Francisella novicida* Cpf1 (FnCpf1), when altered to alanine, did not impact on its DNase capacity while inhibiting RNase activity[181]. Therefore, it is unlikely that the modified residues in dMAD7 would impact on RNase activity, if present. The elucidation of the crystal structure for MAD7, bound to its DNA target, or pre-crRNA transcript, will yield greater insights into other residues for modification, as well as inform experiments where effector proteins, such as deaminases, can be fused to dMAD7 thus increasing the toolbox of MAD7 and dMAD7 based tools.

MAD7 and dMAD7 represent a potential game-changer within the commercial genome engineering community. Where previously a potentially costly licence or royalty fee was required, commercial enterprises without significant budgets now have the ability to harness the revolutionary impact of CRISPR tools. However, it is likely that there remains a wariness to the use of CRISPR-MAD7 for commercial strain construction due to the complexity and rapidly changing IP landscape around CRISPR tools. Commercial enterprises may still wish to proceed with a costly freedom-to-operate search to have full confidence that strains which have been modified with MAD7 will not incur read-through royalty fees from other CRISPR patents. This is likely to remain unclear until a final legal decision is reached on the ownership of the initial CRISPR patents submitted and contest by the University of California and the Broad Institute. Furthermore, companies may not wish to use CRISPR-MAD7 genome editing before the MAD7 patent held by Inscripta, or their claim it is free-to-use for commercial research and strain construction, is unsuccessfully challenged by the eventual owner of the original CRISPR patents. That being said, the CRISPR IP battle is primarily focused on the use of CRISPR-Cas9 tools within eukaryotic hosts. As such, the use of CRISPR tools within bacteria, and the use of Cpf1 based tools, of which MAD7 is a distant relative, could be considered as a less complex research approach upon which to obtain legal clarification in comparison to Cas9 based tools and eukaryotic hosts.

The development of CRISPRa within bacterial hosts has remained rare since its first exemplification within *E. coli* by Bikard *et al*. where dCas9 was fused to the ω subunit of RNAP[59]. Alternative CRISPRa tools have since been established for use within *E. coli* and are based around the SoxS transcriptional regulator[168]*.* However these have since been shown not to be functional within *B. subtilis*[78]. Alternatively, σ^{54} promoters have been successfully targeted in *E. coli* with a integration host factor (IHF) dependent system[170]. This may be transferrable to *B. subtilis,* however it is likely that an alternative to the DNA bending IHF would be required, such as the histone-like protein HBsu, as IHF is only found in Gram-negative bacteria. To date, the only example of CRISPRa in *B. subtilis* was exemplified by Lu *et al*. where dCas9 was fused to the ω or α subunits of RNAP to increase the expression of the extracellular chaperone, PrsA, while simultaneously down regulating (CRISPRi) the extracellular proteases Vpr, Bpr and NprB[60]. In this study, the ω subunit of RNAP was also fused to dCas9, but did not allow the exemplification of CRISPRa under the tested conditions. It is possible that our analysis did not have a sufficiently flexible linking region between the ω subunit and dCas9 (Ala-Ala) to enable interactions between the fusion protein and RNAP in comparison to that used by Lu *et al*. (Gly-Ser-Ala-Ala-Ser). It would be of interest to determine if the linker and effector protein, exemplified by Lu *et al*., could be fused with the dMAD7 developed here for transcriptional upregulation within *B. subtilis*.

Our system has also enabled the expansion of the *B. subtilis* CRISPR toolbox through the development of genome editing tools which do not rely on a DNA DSB. C base editors (CBE) and A base editors (ABE) were shown to function well in *B. subtilis.* Furthermore, the N-terminal fusion of the ABE protein TadA to nCas9 enabled greater selection and isolation of successfully edited cells. Such deaminase base editing could prove a useful tool for extensive multiplexed editing of *B. subtilis* for the introduction of stop codons to disrupt gene expression with CBEs, or the editing of, for example, an enzyme binding pocket with CBE or ABE systems. Very recently, an alternative CRISPR based method was developed in *S. cerevisiae* and human cells which may be worthy of further investigation within bacterial hosts. This prime editing approach utilised nCas9 fused to a reverse transcriptase, alongside an extended sgRNA containing a targeted insertion[182]. However, as this tool requires the use of nCas9, which we have observed to be lethal to *B. subtilis* in the absence of the TadA effector protein, it may not be applicable within this host in its current form. The fusion of an additional ligase to the complex with a flexible

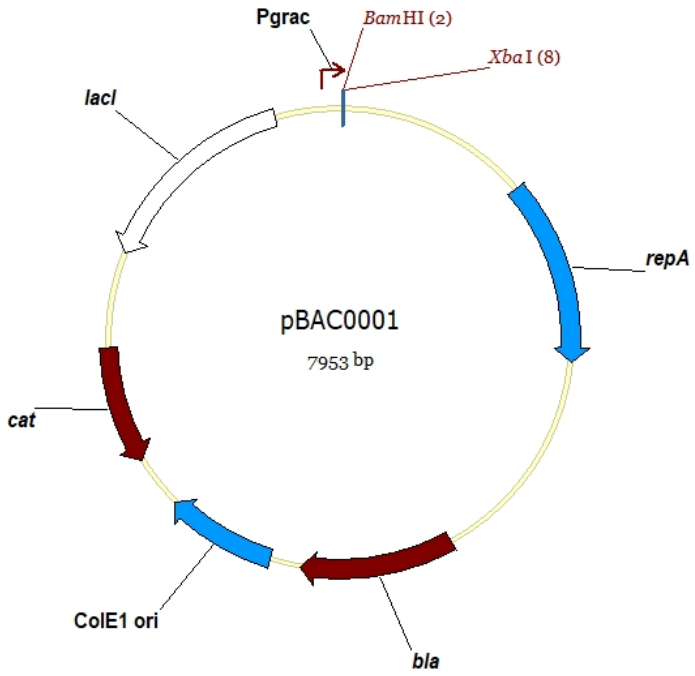
linker may enable repair of the nicked strand and allow prime editing within *B. subtilis*, if the reverse transcriptase alone does not mask the activity of nCas9 in a similar manner to TadA.

In conclusion, this study has advanced the CRISPR toolbox available for use within *B. subtilis*. The utilisation of MAD7 will enable commercial enterprises to develop novel strains in a more rapid manner than previously, while dMAD7 can aid in the identification of genes to be deleted, or their expression lowered. Academically, there are many avenues of research open where further CRISPR tools can be developed and investigated, as outlined above, for use of these tools within *B. subtilis* or other organisms. We hope the tools developed here are taken up for common use within the scientific community for research and commercial purposes.

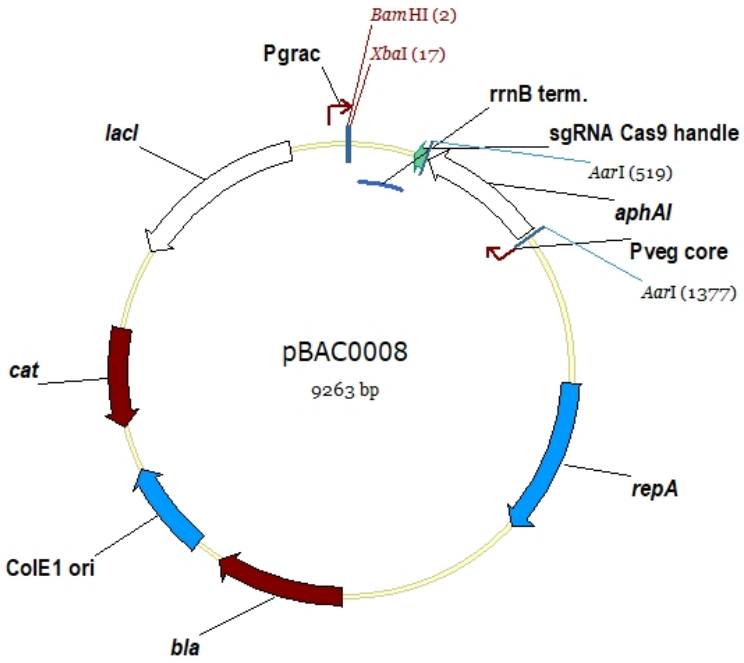
Appendix A

Schematic diagram of the major plasmids used in this study.

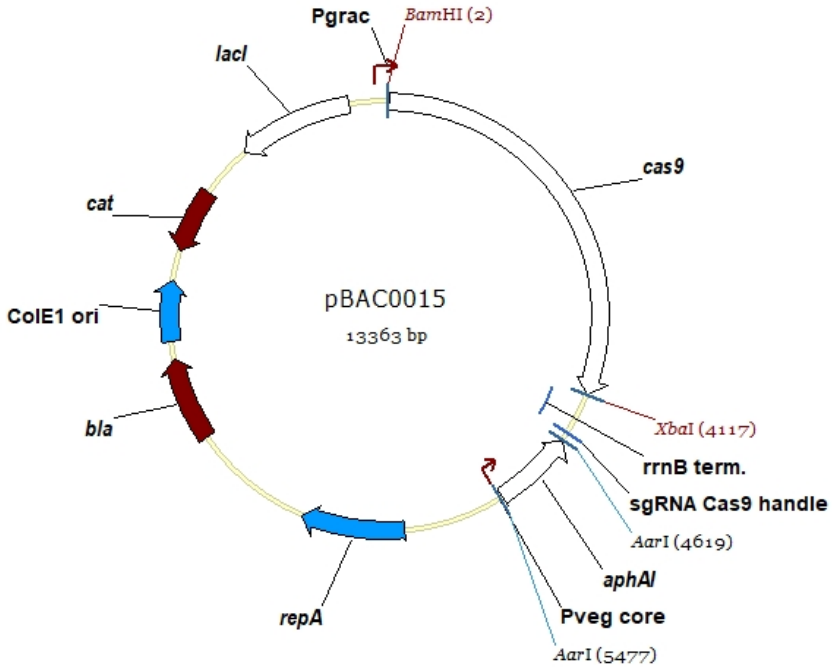
pBAC0001



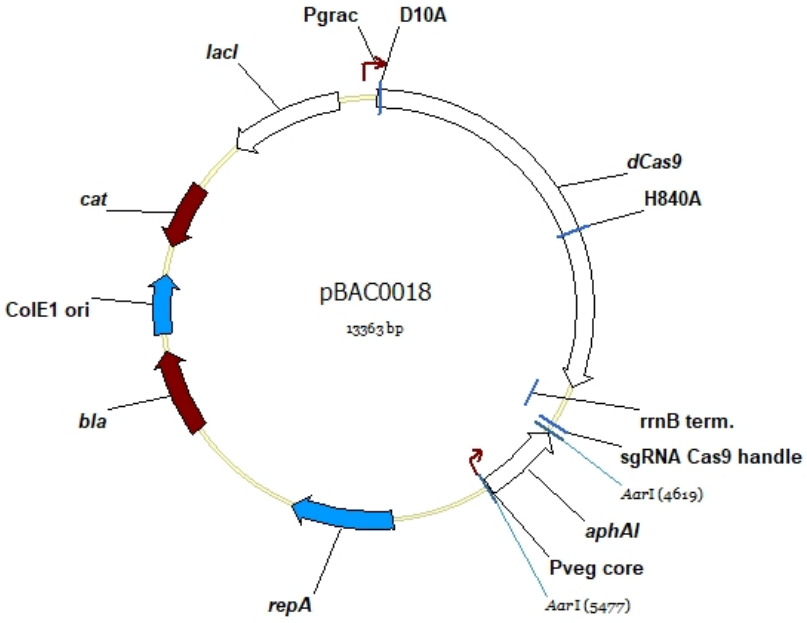
pBAC0008



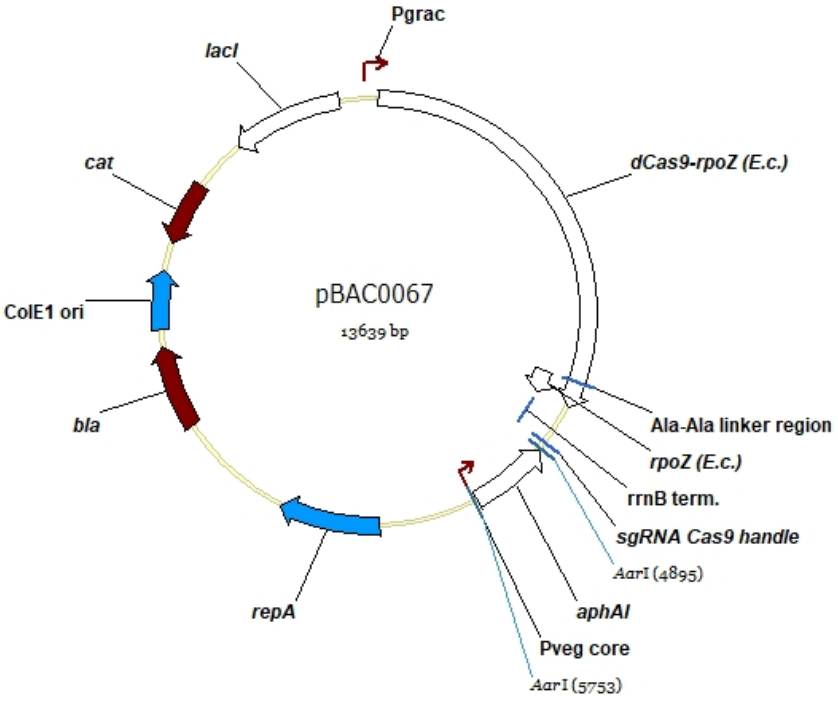
pBAC0015



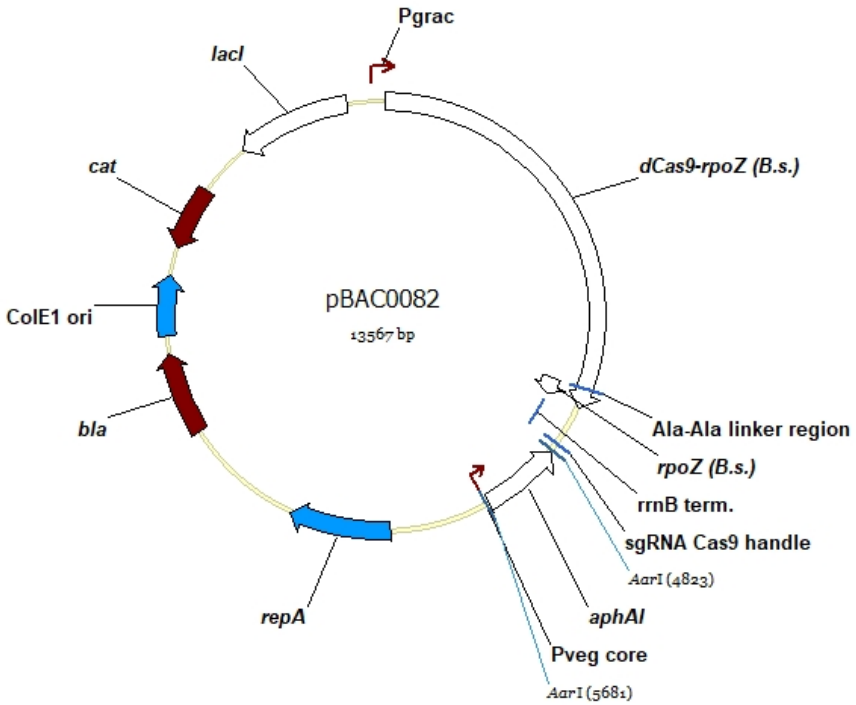
pBAC0018



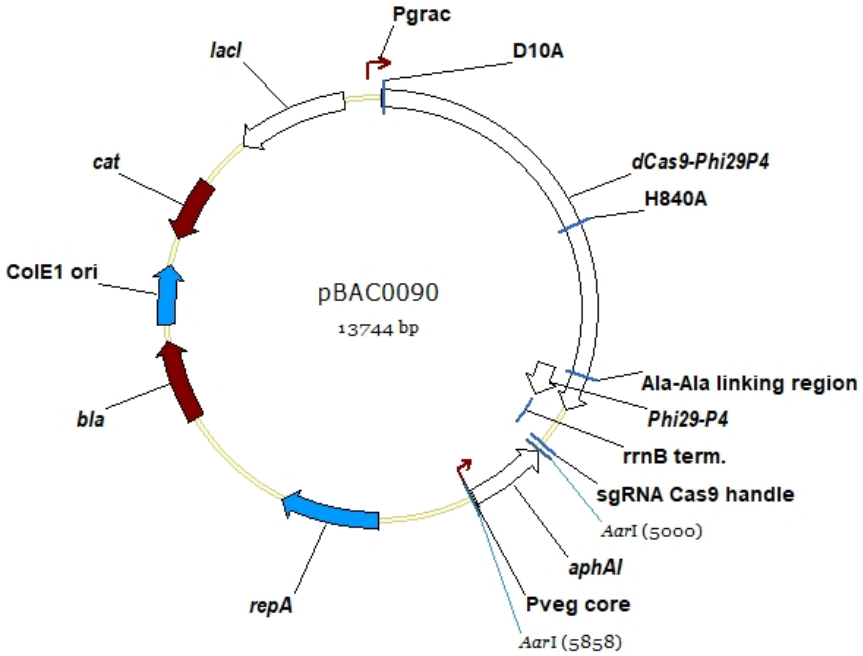
pBAC0067



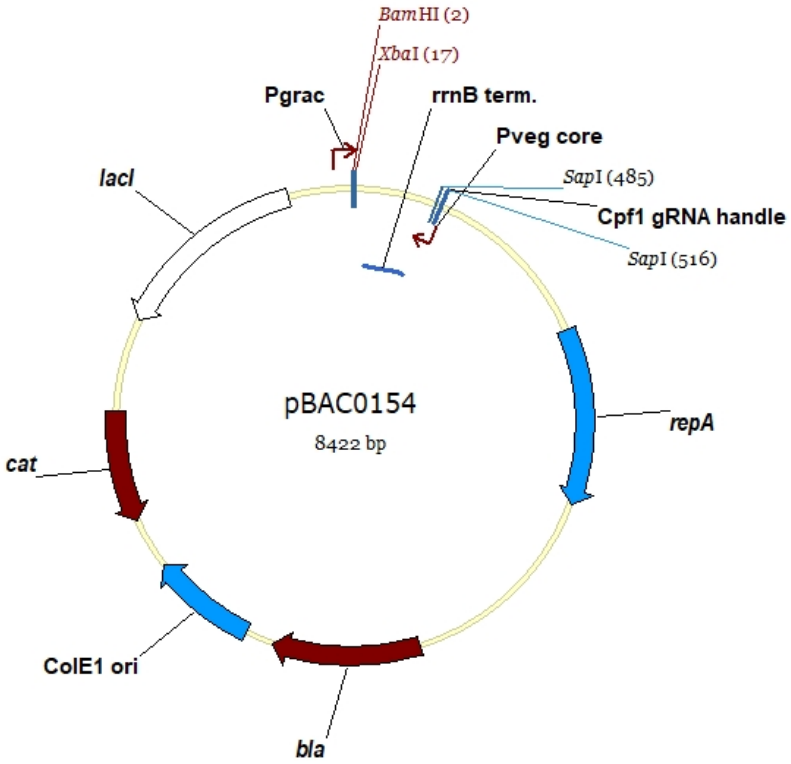
pBAC0082



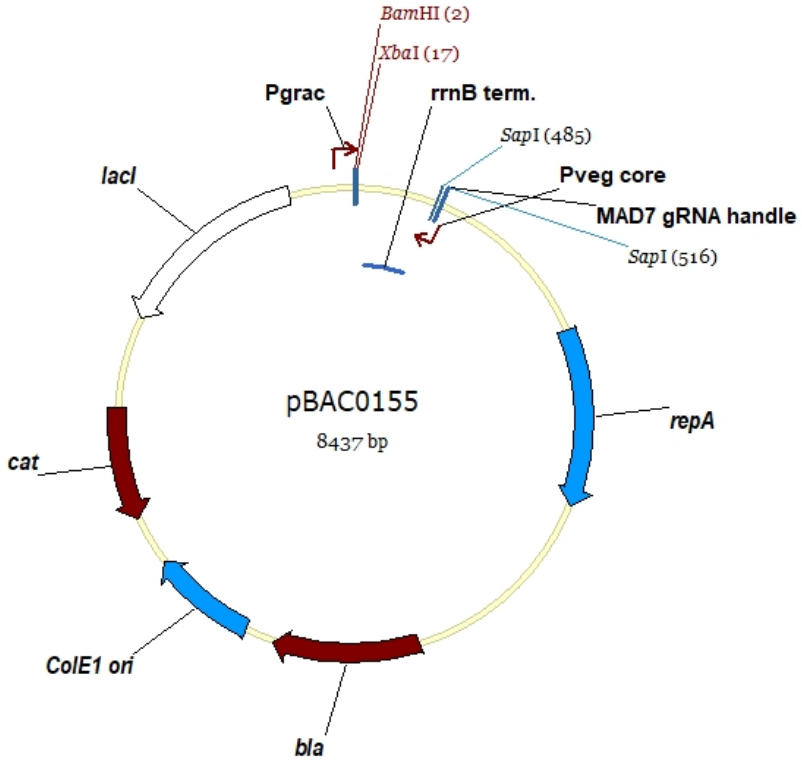
pBAC0090



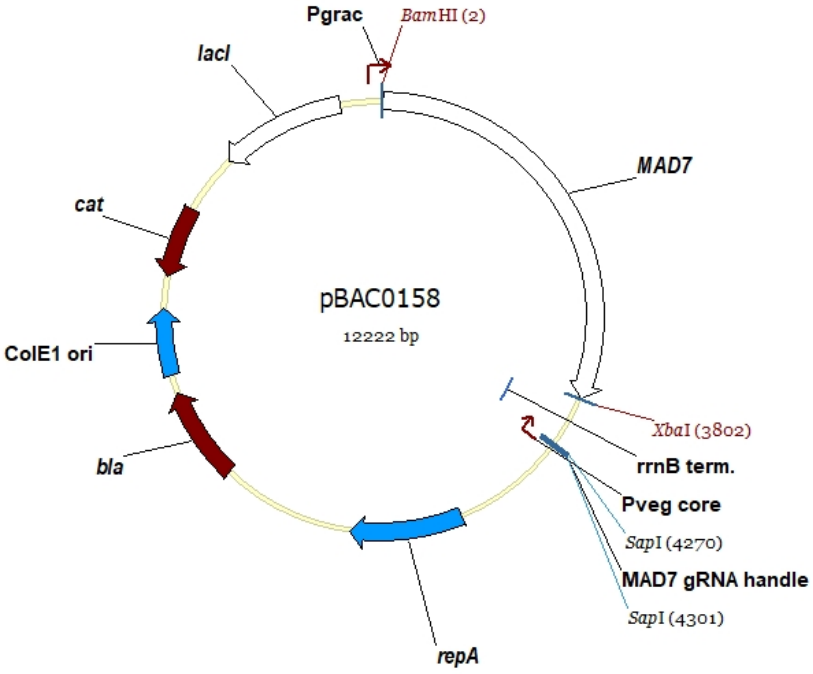
pBAC0154



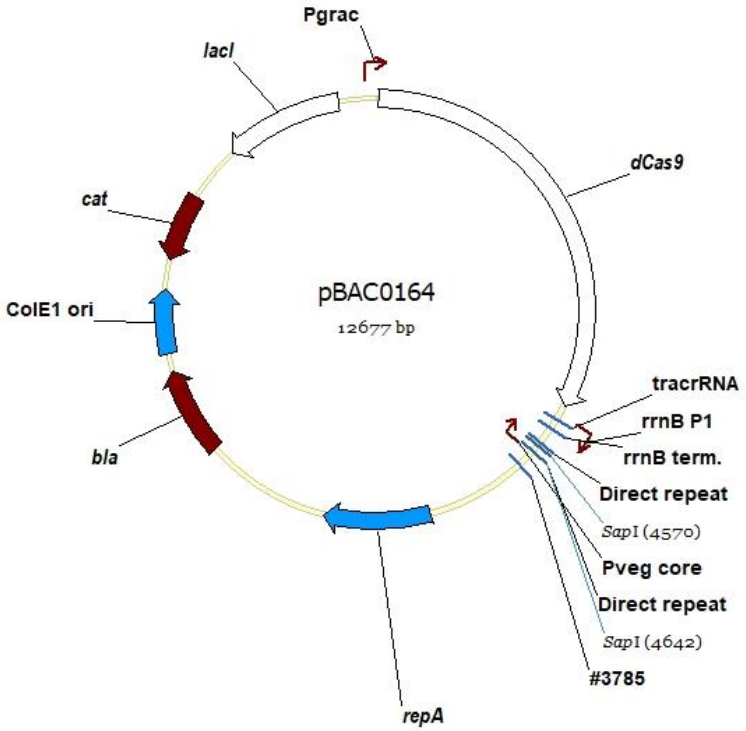
pBAC0155



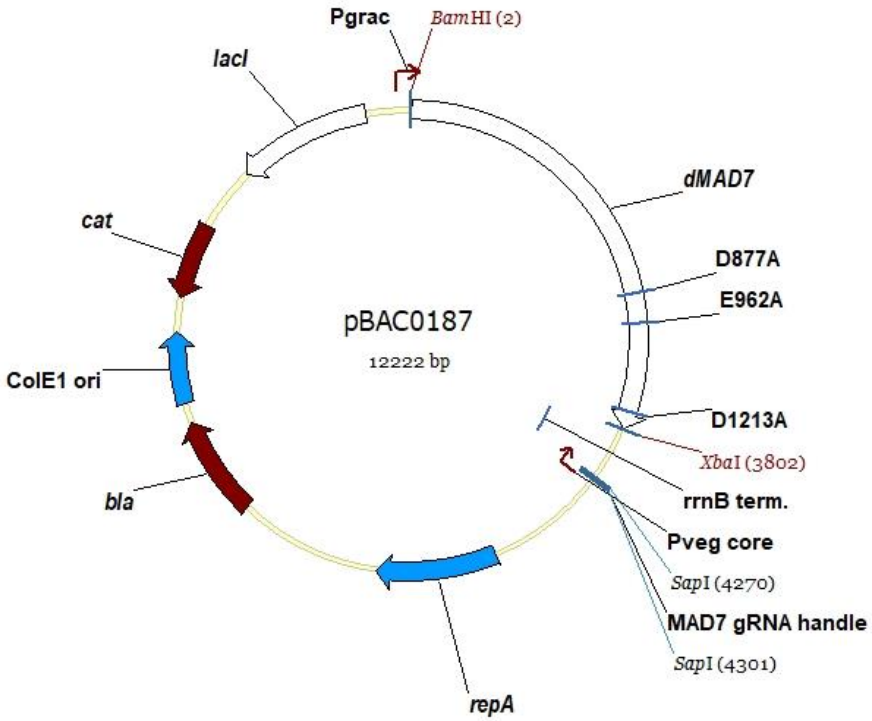
pBAC0158



pBAC0164



pBAC0187



Appendix B

Growth Media used in this study.

LB

Tryptone	10 g/L
Sodium chloride	10 g/L
Yeast Extract	5 g/L
pH	7.2-7.5

LB agar

Tryptone	10 g/L
Sodium Chloride	10 g/L
Yeast Extract	5 g/L
Agar	15 g/L
pH	7.2-7.5

No salt LB

Tryptone	10 g/L
Yeast Extract	5 g/L
pH	7.2-7.5

CHG

Ammonium chloride	1.34 g/L
Ammonium nitrate	0.1 g/L
Potassium dihydrogen phosphate	1.36 g/L
Sodium sulphate	0.11 g/L
L-glutamic acid	3.68 g/L
L-asparagine	1.39 g/L
L-alanine	1.2 g/L
Casein hydrolysate	10 g/L
Ferric chloride hexahydrate	0.001 g/L
Magnesium sulphate heptahydrate	0.99 g/L
Calcium chloride	0.02 g/L
Manganese(II) sulphate tetrahydrate	0.022 g/L
L-tryptophan	0.02 g/L
Glucose	5 g/L
pH	7

M9 agar

Sodium hydrogen phosphate heptahydrate	12.8 g/L
Potassium dihydrogen phosphate	3 g/L
Sodium chloride	0.5 g/L
Ammonium chloride	1 g/L
Magnesium sulphate heptahydrate	0.25 g/L
Glucose	4 g/L
Calcium chloride dihydrate	0.015 g/L
Agar	15 g/L
pH	7.4

Starvation medium 1 (SM1)

Ammonium sulphate	2 g/L
Dipotassium hydrogen phosphate	12 g/L
Potassium dihydrogen phosphate	6 g/L
Sodium citrate dehydrate	1 g/L
Magnesium sulphate heptahydrate	2 g/L
Yeast extract	2 g/L
Casamino acids	0.25 g/L
L-tryptophan	0.1 g/L
Glucose	5 g/L
pH	7

Starvation medium 2 (SM2)

Ammonium sulphate	2 g/L
Dipotassium hydrogen phosphate	12 g/L
Potassium dihydrogen phosphate	6 g/L
Sodium citrate dehydrate	1 g/L
Magnesium sulphate heptahydrate	8 g/L
Yeast extract	1 g/L
Casamino acids	0.125 g/L
L-tryptophan	0.1 g/L
Glucose	5 g/L
Calcium chloride dihydrate	0.22 g/L
pH	7

Appendix C

Oligonucleotides used in this study.

Oligonucleotide Name	Sequence (5´-3')
oMAP0002	tattgggcgctctaccgcttcc
oMAP0003	gatatggacagaggagcaaa
oMAP0004	tttgctcctctgtccatatc
oMAP0005	aagcggtagagcgcccaatacgcat
oMAP0008	ctattcctaataagccgatattagcctcgtatgtt
oMAP0009	atcggcttattaggaa
oMAP0010	gaggaagcggccgcgctcttcgtagtccctgtactaataaaatcagg
oMAP0011	ttagatggatccgctcttcgggctggccgtcgttttacaac
oMAP0012	gccgctagcataatacctaggactgagctagctgtaaatggaacaaatggttgaaacatacgaggctaat
oMAP0013	tcaaccatttgttccatttacagctagctcagtcctaggtattatgctagc
oMAP0014	gccgtgaattcgagctcaggccttaactcacatt
oMAP0015	tgagctcgaattcac
oMAP0016	cagtgagcgcaacgcaattaatgtgagttaaggcc
oMAP0017	aattgcgttgcgctca
oMAP0018	gaggaagcggccgcgctcttcgctgcccgctttccag
oMAP0019	ttagataagcttgctcttcgggcggatccttcctcctttaattgg
oMAP0020	gccgactgctctagagtcgacgtccccggggcagc
oMAP0021	cgactctagagcagtc
oMAP0022	atctccatggacgcgtgacgtgaaaaaagcccgctcattaggcgggctgccccggggacgt
oMAP0023	ccgcctaatgagcgggctttttttcacgtcacgcgtccatgga
oMAP0024	gaggaagcggccgcgctcttcagatgtttgtagaaacgcaaaaaggccatccg
oMAP0025	ttagataagcttgctcttcgggcaacagaatttgcctggcg
oMAP0030	gaggaagcggccgcgctcttcgacggcaggtgttagaaaaactcatcgagc
oMAP0031	ttagataagcttgctcttcgggcagcaggtgtaaggaggcatcgtgaatgagccatattcaacgggaaacg
oMAP0048	gccaaaaaaagcaccgactcggtgccacttttttcaag
oMAP0049	cgagtcggtgctttttt
oMAP0050	cgtagttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcac
oMAP0051	ttgataacggactagccttattttaacttgctatttctagctctaaaact
oMAP0052	gccacatttattgtacaacacgagcccatttttgtcaacaacactggaacaaatggttgaaacatacgaggctaat
oMAP0053	tcaaccatttgttccagtgttgttgacaaaaatgggctcgtgttgtacaataaatgt
oMAP0054	gccgtttaataatataacacctgtatttatttattgcaatagttttggaacaaatggttgaaacatacgaggctaat
oMAP0055	tcaaccatttgttccaaaactattgcaataaataaatacaggtgttatattattaaac
oMAP0062	tactcaataggcttagatatcggcacaaatagcgtcgg
oMAP0063	acgctatttgtgccgatatctaagcctattgag
oMAP0064	gtgattatgatgtcgatcacattgttccacaaagtttcc
oMAP0065	actttgtggaacaatgtgatcgacatcataatcac
oMAP0066	cgtaaaaaagactgggacccaaaaaaatatggtgg
oMAP0067	ccaccatatttttttgggtcccagtctttttacg
oMAP0073	cactgaggtctctgatcctcatggataagaaatactcaataggc
oMAP0074	tcgtactctagattagtcacctcctagctgactcaaatcaatgc
oMAP0089	aaacctttgatcgacggtgctgt
oMAP0091	atgtacagcaccgtcgatcaaaag
oMAP0121	atccggcgttctcatggcgg
oMAP0122	tattgaccgcagtgatagcc

oMAP0123	atgcatgaagaatggttcagcttttgatcgacggtgctgt
oMAP0124	acagcaccgtcgatcaaaagctgaaccattcttcatgcatgg
oMAP0125	aaacctttactgccgttattcgct
oMAP0127	atgtagcgaataacggcagtaaag
oMAP0128	aataacggcagtaaagttaagttttgaatcgttttgc
oMAP0129	aaacgattcaaaacttaactttactgccgttattcgctgg
oMAP0140	aaacgtcctgccagaaccaaatga
oMAP0142	atgttcatttggttctggcaggac
oMAP0143	gtttcagcactcgcagcctacggtcctgccagaaccaaatg
oMAP0144	catttggttctggcaggaccgtaggctgcgagtgctgaaac
oMAP0145	aaaccgagtcagctagctagcatg
oMAP0147	atgtcatgctagctagctgactcg
oMAP0148	atcccctaattttcgtaccattaaccttccacc
oMAP0150	aaacgcgctttaatttgagaaatg
oMAP0151	atgtcatttctcaaattaaagcgc
oMAP0152	aagaaaacgttggatagagc
oMAP0153	tcgagagtgaagagcaggcgctttaatttgagaaatgcc
oMAP0154	aaagcgcctgctcttcactctcgaggctacacaggctctaacg
oMAP0155	ttgacagagaacagagaagc
oMAP0156	aaacgcttctttttactattattg
oMAP0157	atgtcaataatagtaaaaagaagc
oMAP0158	ttaaacagacaatgagtgcc
oMAP0159	gcttctttttactattatcgtgcagctgcttctacgttgattaacccttttcc
oMAP0160	cgataatagtaaaaagaagcagtttcctccatacctgcttc
oMAP0161	tttcgctgattacaacattgg
oMAP0186	gatcgtcccgggttattgtgcagctgcttgtacg
oMAP0195	atggactgaaacatcatcggc
oMAP0198	ttataagcgccgtaagtgcc
oMAP0199	cagatcggatccgagtgagaagcaaaaaattgtgg
oMAP0200	gatcgtcccgggttatcgtgcagctgcttctacg
oMAP0203	gaggaaggatccgaatgagaagcaaaaaattgtgg
oMAP0208	aaacacatatttgtgtgaaatcgt
oMAP0209	atgtacgatttcacacaaatatgt
oMAP0210	tacagacggtcttactaggc
oMAP0211	tgattattcaacctccagc
oMAP0212	aaatgctggaggttgaataatcatagtagcacaagtagcaacctatatcatgtagg
oMAP0213	ttcagttctaaggtcatgcc
oMAP0214	gtcacctcctagctgactcaaatcaatgc
oMAP0215	tttgagtcagctaggaggtgacgcagctttagatccgtcaattgattctttaatg
oMAP0216	gtcgactctagactattcgcggtcttccttttcaaacg
oMAP0217	tacaacattggtgacgctgcc
oMAP0218	gctaagcagaaggccatcctgtagtaaaaagaagcaggttcctccatacctgc
oMAP0219	cgttcaggcagtacgcctcttttttctttaccctctccttttaaaaaaattcagag
oMAP0220	atcggcatttttttctgcctgc
oMAP0221	caggatggccttctgcttagc
oMAP0222	aaaagaggcgtactgcctgaacg
oMAP0223	aaatggcacgtgaaaatcagac
oMAP0224	aaaccactcattgcaatttaggtt
oMAP0225	atgtaacctaaattgcaatgagtg
oMAP0226	aaactaggcgtgagtaacgttaaa
oMAP0227	atgttttaacgttactcacgccta
oMAP0228	aaacgctttatcagtaacaatcct
oMAP0229	atgtaggattgttactgataaagc
oMAP0230	aaacttttcatatgtgaatgtaag
oMAP0231	atgtcttacattcacatatgaaaa

oMAP0232	aaacagattgtttttcacttgtgt
oMAP0233	atgtacacaagtgaaaaacaatct
oMAP0244	ccgcgggctttcccagccaaaaatcagaccagacaaaagcggc
oMAP0245	gttcctccttcccacctcattcattctattataaaggaaaagc
oMAP0252	tcaaaaatcccgccattgcc
oMAP0255	atctaatgagccggaaagcgg
oMAP0258	agcagttcatattttctcaggg
oMAP0261	tagatcaagcatgctgacaccg
oMAP0264	atatctgaaacgcgagcgg
oMAP0267	aatcggctcgacatcattgc
oMAP0317	tttgagtcagctaggaggtgacg
oMAP0318	acgtcgactctagattacacc
oMAP0320	aaacgcggtaaaagaattcctgca
oMAP0321	atgttgcaggaattcttttaccgc
oMAP0322	aaacgcagtcctgtaactatcatc
oMAP0323	atgtgatgatagttacaggactgc
oMAP0324	aaacgtaactatcatcatcttccc
oMAP0325	atgtgggaagatgatgatagttac
oMAP0326	aaactctgaaaaaaataaaaacct
oMAP0327	atgtaggtttttatttttttcaga
oMAP0328	aaacaaattaccatcagcacagac
oMAP0329	atgtgtctgtgctgatggtaattt
oMAP0330	aaactcagcacagacagagatgat
oMAP0331	atgtatcatctctgtctgtgctga
oMAP0336	aaaccgggctttcccagccaaaaa
oMAP0337	atgtttttggctgggaaagcccg
oMAP0338	aaacagccaaaaatcagaccagac
oMAP0339	atgtgtctggtctgatttttggct
oMAP0340	aaacaaaatcagaccagacaaaag
oMAP0341	atgtcttttgtctggtctgatttt
oMAP0342	aaacgacaaaagcggcaaatgaat
oMAP0343	atgtattcatttgccgcttttgtc
oMAP0344	aaacctttgtctggtctgatttt
oMAP0345	atgtaaaatcagaccagacaaaag
oMAP0346	aaaccttattcatttgccgctttt
oMAP0347	atgtaaaagcggcaaatgaataag
oMAP0350	aaaccgttccgcttattcatttgc
oMAP0351	atgtgcaaatgaataagcggaacg
oMAP0354	aaaccaaatccttccccgttccgc
oMAP0355	atgtgcggaacggggaaggatttg
oMAP0386	aaacgggaaagcccgcggtaaaag
oMAP0387	atgtcttttaccgcgggctttccc
oMAP0388	tcttcatcctcttcgtcttgg
oMAP0393	aaagaattggtcgtcatggg
oMAP0394	aaatgggctcgtgttgtacaataaatgtgggaaggaggaactactatgcg
oMAP0395	tgtacaacacgagcccatttttgtcaacaacacgggaaagcccgcggtaaaagaattcc
oMAP0396	ataaataaatacaggtgttatattattaaacgggaaggaggaactactatgcg
oMAP0397	taacacctgtatttatttattgcaatagtttgggaaagcccgcggtaaaagaattcc
oMAP0406	aaacatttttgtcaacaacacggg
oMAP0407	atgtcccgtgttgttgacaaaaat
oMAP0412	aaacgtatttatttattgcaatag
oMAP0413	atgtctattgcaataaataaatac
oMAP0414	aaacaaactattgcaataaataaa
oMAP0415	atgttttatttattgcaatagttt
oMAP0486	gcccaaaaaaacccctcaagacccgtttagaggccccaaggggttatgctagg

oMAP0487	ggggcctctaaacgggtcttgaggggtttttg
oMAP0488	tacgtagctgctcttccctagcataacccctt
oMAP0489	gaagagcagctac
oMAP0490	gtacgtgctcttccatctacaagagtagaaatta
oMAP0491	tgttaatttctactcttgtagatggaagagcacg
oMAP0492	gtacgtgctcttccatctacaagagtagaaattaaaaaggtcttttgac
oMAP0493	tgtgtcaaaagacctttttaatttctactcttgtagatggaagagcacg
oMAP0498	acatttattgtacaacacgagcccatttttgtcaacaacactggaacaaatggttgaaacatacgaggctaat
oMAP0499	tcaaccatttgttccagtgttgttgacaaaaatgggctcgtgttgtacaataaa
oMAP0500	acgtagggatcctcatgacacagttcgagggctttacc
oMAP0501	acgtagtctagattagttgcgcagctcctggatgtaggcc
oMAP0547	taggtcaagaatgtttgcaaaacgatt
oMAP0548	gataatcgttttgcaaacattcttgac
oMAP0549	tagaagaatgtttgcaaaacgatt
oMAP0550	gataatcgttttgcaaacattctt
oMAP0551	ctattagaatcgttttgcaaacattcttgacactcc
oMAP0552	ttgcaaaacgattctaatagtctttactgccgttattcgctgg
oMAP0553	tagacgtcgagtcagctagctagcatg
oMAP0554	gatcatgctagctagctgactcgacgt
oMAP0555	tagtcgagtcagctagctagcatg
oMAP0556	gatcatgctagctagctgactcga
oMAP0563	aaccgagtcagctagctagcatggttttgggaccattcaaaacagcatagctctaaaacgcggtaaaa
oMAP0564	attcttttaccgcgttttagagctatgctgttttgaatggtcccaaaaccatgctagctagctgactcg
oMAP0565	gaattcctgcagttttgggaccattcaaaacagcatagctctaaaacgcgaataacggcagtaaaga
oMAP0566	aactctttactgccgttattcgcgttttagagctatgctgttttgaatggtcccaaaactgcagga
oMAP0569	tagaacaagaattgggacaactccagt
oMAP0570	gatactggagttgtcccaattcttgtt
oMAP0571	tagaagaattgggacaactccagt
oMAP0572	gatactggagttgtcccaattctt
oMAP0573	aaacgtgaaaagttcttctccttt
oMAP0574	atgtaaaggagaagaacttttcac
oMAP0575	tgttcacgctcgataacaag
oMAP0578	aatctgcctcctcatcctcttcatcc
oMAP0638	gtgggtctcggatcctcatgagcgaggtggaattcagcc
oMAP0639	tggggtctccctagattagtcacctcctagctgactc
oMAP0640	gtgggtctcggatcctcatggataagaaatactcaataggc
oMAP0641	tggggtctccctagattagccgctaccacggctaaccg
oMAP0670	caccatccttcggcaaatcc
oMAP0671	attgcgggaaatgcagtggc
oMAP0682	aaacccctgttagcgtgggggggt
oMAP0683	atgtacccccccacgctaacaggg
oMAP0694	aaacctccgattcttcttgattcc
oMAP0695	atgtggaatcaagaagaatcggag
oMAP0701	taggagagggtaaagagtgagaag
oMAP0702	gatcttctcactctttaccctctc
oMAP0703	gtgcagaagctagaaattgtggatcagcttgttgtttgc
oMAP0704	ccacaatttctagcttctgcactctttaccctctccttttaaaa
oMAP0705	tagagcaacagacaatttcttacc
oMAP0706	gatggtaagaaattgtctgttgct
oMAP0707	gtgtgagtgaggtaagaaattgtctgttgctgtcg

oMAP0708	caatttcttacctcactcacacataaatccccctttttgaaaatac
oMAP0709	tagatgatggataactatcgtgat
oMAP0710	gatatcacgatagttatccatcat
oMAP0711	atgtagataactatcgtgattagaacagaacgaaaggtaatgag
oMAP0712	tctaatcacgatagttatctacatcatgttcactccgtttctctatttt
oMAP0713	tagatggattatcgacgtgatggc
oMAP0714	gatgccatcacgtcgataatccat
oMAP0715	atgtgattatcgacgtgatggctagaacgatcaacatcaaactgaacc
oMAP0716	ctagccatcacgtcgataatcacatgttcttacactcctttaacgg
oMAP0717	tagtacaaaggggaaacacattg
oMAP0718	gatcaatgtgtttccccctttgta
oMAP0719	gcacgacaaagatcttttcc
oMAP0720	atctacaatgtgtttccccctttgtatttag
oMAP0721	aaaggggaaacacattgtagataggggatcattcgctttctgc
oMAP0722	tccgctgttgccgtttgagg
oMAP0724	taggcatgttgaaaaaggggatg
oMAP0725	gatcatccccctttttcaacatgc
oMAP0726	tctcggcaatgaaagaagcg
oMAP0727	tttttcgttttttttcctacatattcatcccccttttttcaacatgc
oMAP0728	aatatgtaggaaaaaaacgaaaaacagactcatcagc
oMAP0729	tgctaaaacccattcaccagc
oMAP0731	taggatgtcttggtcaagttgcgc
oMAP0732	gatgcgcaacttgaccaagacatc
oMAP0733	aatgcagagctggcgaatatgcc
oMAP0734	ttcacaagacaccacatccttcctattttgg
oMAP0735	aaggatgtggtgtcttgtgaacttgaccaagacatctctattactgg
oMAP0736	atgcacggtagacgtaatcc
oMAP0738	taggacactaaaggagggagatga
oMAP0739	gattcatctccctcctttagtgtc
oMAP0740	caagctgacgatggatcggg
oMAP0741	aatcttggaactaattcatcctgtcatctccctcctttagtgtc
oMAP0742	caggatgaattagttccaagattcagaaaacaatgg
oMAP0743	tagacatgaatatggctagtttgc
oMAP0745	tagtagaaaaaggagatgaatcat
oMAP0746	gatatgattcatctccttttttcta
oMAP0747	ttccgccaatgatggaagcc
oMAP0748	gtttgcaagacatgttcataacatgattcatctccttttttctatga
oMAP0749	atgttatgaacatgtcttgcaaacttgttgtatcag
oMAP0750	ttgggatgttccgcttcccg
oMAP0752	taggagggataacatgaaacgcag
oMAP0753	gatctgcgtttcatgttatccctc
oMAP0754	aaggcataagcagtttttatcgg
oMAP0755	agctgaatcatctgcgttttacatgttatccctcctgcaaaataatg

oMAP0756	atgtaaaacgcagatgattcagctcggttgtggcggc
oMAP0757	ttgtccttcaccttgaaaataacc
oMAP0759	tagggaaggaggaactactatgcg
oMAP0760	gatcgcatagtagttcctccttcc
oMAP0761	tagggaactactatgcgtaaagga
oMAP0762	gattcctttacgcatagtagttcc
oMAP0763	tagtaacatcaccatctaattcaa
oMAP0764	gatttgaattagatggtgatgtta
oMAP0765	tagagatggtgatgttaatgggca
oMAP0766	gattgcccattaacatcaccatct
oMAP0767	tagccttcaccctctccactgaca
oMAP0768	gattgtcagtggagagggtgaagg
oMAP0769	tagaagaggttttgaatcgttttg
oMAP0770	gatcaaaacgattcaaaacctctt
oMAP0771	tagaaggagtgtcaagaatgtttg
oMAP0772	gatcaaacattcttgacactcctt
oMAP0775	tagaaatccagcgaataacggcag
oMAP0776	gatctgccgttattcgctggattt
oMAP0777	tagcagaaccaaatgaaacagcaa
oMAP0778	gatttgctgtttcatttggttctg
oMAP0793	tagatctacaagagtagaaattaaaaaggtcttttgacaagaattgggacaactccagtatctaca agagtagaaattaaaaaggtcttttgacccttcaccctctccactgaca
oMAP0794	gattgtcagtggagagggtgaagggtcaaaagacctttttaatttctactcttgtagatactggagtt gtcccaattcttgtcaaaagacctttttaatttctactcttgtagat
oMAP0795	tagatctacaagagtagaaattaaaaaggtcttttgaccagaaccaaatgaaacagcaaatcta caagagtagaaattaaaaaggtcttttgacaagaggttttgaatcgttttg
oMAP0796	gatcaaaacgattcaaaacctcttgtcaaaagacctttttaatttctactcttgtagatttgctgtttcattt ggttctggtcaaaagacctttttaatttctactcttgtagat
oMAP0797	tagatctacaagagtagaaattaaaaaggtcttttgacaagaggttttgaatcgttttgatctacaag agtagaaattaaaaaggtcttttgacccttcaccctctccactgaca
oMAP0798	gattgtcagtggagagggtgaagggtcaaaagacctttttaatttctactcttgtagatcaaaacgat tcaaaacctcttgtcaaaagacctttttaatttctactcttgtagat
oMAP0799	tagtcttttctaagtacaggaatc
oMAP0800	gatgattcctgtacttagaaaaga
oMAP0803	gtcatgcaagagattttgag
oMAP0804	gtgcccattaacatcacc
oMAP0811	ggtcgtcaactgtgtgaacc
oMAP0812	gagagttcctcaggctgtcc
oMAP0813	tgatttgtattcactctgcc
oMAP0814	tgaaatccttacatacggcg
oMAP0815	ctaaattgcaatgagtgcgg

Appendix D

DNA alignment report comparing MAD7 as provided by Inscripta Inc., *Bacillus subtilis* codon optimised MAD7 and dMAD7 sequences. *B. subtilis* codon optimisation was performed by Thermo Fischer Scientific.

```
Summary of Percent Matches:
   Ref:  MAD7 - Inscripta       1 to   3792   (  3792 bps)     --
     2:  MAD7 - B. subtil       1 to   3792   (  3792 bps)     79%
     3:  dMAD7                  1 to   3792   (  3792 bps)     79%


MAD7 - Inscripta      1 atgaacaacgggcacaaatatattttcagaacttcatcgggatctcaagtttgcagaaaacgctgcgcaatgctctgatccc
MAD7 - B. subtil      1 atgaataatggcacaaacaacttttcagaacttcattggcattagcagcctgcaaaaaacactgagaaatgcactgattcc
dMAD7                 1 atgaataatggcacaaacaacttttcagaacttcattggcattagcagcctgcaaaaaacactgagaaatgcactgattcc

MAD7 - Inscripta     81 cacggaaaccacgcaacagttcatcgtcaagaacggaataattaaagaagatgagttacgtggcgagaaccgccagattc
MAD7 - B. subtil     81 gacagaaacaacacagcagtttattgtcaaaaacggcatcatcaaagaggatgaactgagaggcgaaaatcgccaaattc
dMAD7                81 gacagaaacaacacagcagtttattgtcaaaaacggcatcatcaaagaggatgaactgagaggcgaaaatcgccaaattc

MAD7 - Inscripta    161 tgaaagatatcatggatgactactaccgcggattcatctctgagactctgagttctattgatgacatagattggactagc
MAD7 - B. subtil    161 tgaaagatatcatggacgactattaccgtggctttatttcagaaacactgtccagcattgatgatatcgattggacaagc
dMAD7               161 tgaaagatatcatggacgactattaccgtggctttatttcagaaacactgtccagcattgatgatatcgattggacaagc

MAD7 - Inscripta    241 ctgttcgaaaaaatggaaattcagctgaaaaatggtgataataaagataccttaattaaggaacagacagagtatcggaa
MAD7 - B. subtil    241 ctgttcgagaaaatggaaatccaactgaaaaacggcgataacaaagacacgctgattaaagaacaaacggaatatcgcaa
dMAD7               241 ctgttcgagaaaatggaaatccaactgaaaaacggcgataacaaagacacgctgattaaagaacaaacggaatatcgcaa

MAD7 - Inscripta    321 agcaatccataaaaaatttgcgaacgacgatcggtttaagaacatgtttagcgccaaactgattagtgacatattacctg
MAD7 - B. subtil    321 agcgatccacaaaaagtttgcaaatgatgaccgctttaaaaacatgttcagcgcgaaactgattagcgatattctgccgg
dMAD7               321 agcgatccacaaaaagtttgcaaatgatgaccgctttaaaaacatgttcagcgcgaaactgattagcgatattctgccgg

MAD7 - Inscripta    401 aatttgtcatccacaacaataattattcggcatcagagaaagaggaaaaaacccaggtgataaaattgttttcgcgcttt
MAD7 - B. subtil    401 aatttgtcatccacaatataactatagcgcgagcgagaaagaagaaaaaacacaggtcattaaactgtttagccgcttt
dMAD7               401 aatttgtcatccacaataataactatagcgcgagcgagaaagaagaaaaaacacaggtcattaaactgtttagccgcttt

MAD7 - Inscripta    481 gcgactagctttaaagattacttcaagaaccgtgcaaattgcttttcagcggacgatatttcatcaagcagctgccatcg
MAD7 - B. subtil    481 gccacaagcttcaaagactatttcaaaaatcgcgcaaactgctttagcgcagatgatatttcatcatcaagctgccatcg
dMAD7               481 gccacaagcttcaaagactatttcaaaaatcgcgcaaactgctttagcgcagatgatatttcatcatcaagctgccatcg

MAD7 - Inscripta    561 catcgtcaacgacaatgcagagatattcttttcaaatgcgctggtctaccgccggatcgtaaaatcgctgagcaatgacg
MAD7 - B. subtil    561 gattgtcaatgataatgcggaaatcttttttagcaacgcactggtctatcgcagaattgttaaatcattgagcaacgacg
dMAD7               561 gattgtcaatgataatgcggaaatcttttttagcaacgcactggtctatcgcagaattgttaaatcattgagcaacgacg

MAD7 - Inscripta    641 atatcaacaaaatttcgggcgatatgaaagattcattaaaagaaatgagtctggaagaaatatattcttacgagaagtat
MAD7 - B. subtil    641 acatcaacaaaatctcaggcgatatgaaagacagcctgaaagaaatgtcactggaagaaatctacagctacgaaaaatac
dMAD7               641 acatcaacaaaatctcaggcgatatgaaagacagcctgaaagaaatgtcactggaagaaatctacagctacgaaaaatac

MAD7 - Inscripta    721 ggggaatttattacccaggaaggcattagcttctataatgatatctgtgggaaagtgaattcttttatgaacctgtattg
MAD7 - B. subtil    721 ggcgaatttatcacacaagaaggcatcagcttttacaacgatatttgcggcaaagtcaacagctttatgaatctgtattg
dMAD7               721 ggcgaatttatcacacaagaaggcatcagcttttacaacgatatttgcggcaaagtcaacagctttatgaatctgtattg

MAD7 - Inscripta    801 tcagaaaaataaagaaaacaaaaatttatacaaacttcagaaacttcacaaacagattctatgcattgcggacactagct
MAD7 - B. subtil    801 ccagaaaaacaaagaaaacaaaaacctgtataaactgcagaaactgcacaagcagattctgtgcattgcagatacatcat
dMAD7               801 ccagaaaaacaaagaaaacaaaaacctgtataaactgcagaaactgcacaagcagattctgtgcattgcagatacatcat

MAD7 - Inscripta    881 atgaggtcccgtataaatttgaaagtgacgaggaagtgtaccaatcagttaacggcttccttgataacattagcagcaaa
MAD7 - B. subtil    881 atgaagtcccgtacaaatttgagagcgacgaagaagtttatcaaagcgttaatggctttctggataacatcagcagcaaa
dMAD7               881 atgaagtcccgtacaaatttgagagcgacgaagaagtttatcaaagcgttaatggctttctggataacatcagcagcaaa

MAD7 - Inscripta    961 catatagtcgaaagattacgcaaaatcggcgataactataacggctacaacctggataaaatttatatcgtgtccaaatt
MAD7 - B. subtil    961 catattgttgaacgcctgagaaaaattggcgataactataatggctacaacctggacaaaatctacatcgtcagcaaatt
dMAD7               961 catattgttgaacgcctgagaaaaattggcgataactataatggctacaacctggacaaaatctacatcgtcagcaaatt

MAD7 - Inscripta   1041 ttacgagagcgttagccaaaaaacctaccgcgactgggaaacaattaatacccgcccgaaattcattacaataatatct
MAD7 - B. subtil   1041 ttacgaaagcgtcagccaaaaaacatatcgcgattgggaaacaattaatacagcgctggaaattcattataacaacattc
dMAD7              1041 ttacgaaagcgtcagccaaaaaacatatcgcgattgggaaacaattaatacagcgctggaaattcattataacaacattc

MAD7 - Inscripta   1121 tgccgggtaacggtaaaagtaaagccgacaaagtaaaaaaagcggttaagaatgatttacagaaatccatcaccgaaata
MAD7 - B. subtil   1121 tgcctggcaacggcaaaagcaaagcagataaagttaaaaaggcggtcaaaaatgacctgcagaaaagcattacagaaatc
dMAD7              1121 tgcctggcaacggcaaaagcaaagcagataaagttaaaaaggcggtcaaaaatgacctgcagaaaagcattacagaaatc

MAD7 - Inscripta   1201 aatgaactagtgtcaaactataagctgtgcagtgacgacaacatcaaagcggagacttatatacatgagattagccatat
MAD7 - B. subtil   1201 aatgaactggtcagcaactacaaactgtgctcagatgataatatcaaggcggaaacgtacatccatgaaattagccatat
dMAD7              1201 aatgaactggtcagcaactacaaactgtgctcagatgataatatcaaggcggaaacgtacatccatgaaattagccatat

MAD7 - Inscripta   1281 cttgaataactttgaagcacaggaattgaaatacaatccggaaattcacctagttgaatccgagctcaaagcgagtgagc
MAD7 - B. subtil   1281 cctgaacaactttgaagcgcaagaactgaaatataacccggaaatccatctggttgaaagcgaactgaaagcaagcgagc
dMAD7              1281 cctgaacaactttgaagcgcaagaactgaaatataacccggaaatccatctggttgaaagcgaactgaaagcaagcgagc

MAD7 - Inscripta   1361 ttaaaaacgtgctggacgtgatcatgaatgcgtttcattggtgttcggtttttatgactgaggaacttgttgataaagac
MAD7 - B. subtil   1361 tgaaaaatgttctggatgtcattatgaacgcgtttcattggtgcagcgtttttatgacagaagaactggtcgataaggat
dMAD7              1361 tgaaaaatgttctggatgtcattatgaacgcgtttcattggtgcagcgtttttatgacagaagaactggtcgataaggat
```

```
MAD7 - Inscripta   1441 aacaatttttatgcggaactggaggagatttacgatgaaatttatccagtaattagtctgtacaacctggttcgtaacta
MAD7 - B. subtil   1441 aacaactttttatgcagaactggaagagatttacgacgaaatttatccggtcatcagcctgtataatctggttcgcaatta
dMAD7              1441 aacaactttttatgcagaactggaagagatttacgacgaaatttatccggtcatcagcctgtataatctggttcgcaatta

MAD7 - Inscripta   1521 cgttacccagaaaccgtacagcacgaaaaagattaaattgaactttggaataccgacgttagcagacggttggtcaaagt
MAD7 - B. subtil   1521 tgtcacacagaaaccgtatagcacgaagaaaatcaaactgaactttggcattccgacactggcagatggctggtcaaaat
dMAD7              1521 tgtcacacagaaaccgtatagcacgaagaaaatcaaactgaactttggcattccgacactggcagatggctggtcaaaat

MAD7 - Inscripta   1601 ccaaagagtattctaataacgctatcatactgatgcgcgacaatctgtattatctgggcatctttaatgcgaagaataaa
MAD7 - B. subtil   1601 caaaagaatatagcaacaacgcgatcatcctgatgcgcgataatcttattatctgggcatttttcaacgcgaaaaacaag
dMAD7              1601 caaaagaatatagcaacaacgcgatcatcctgatgcgcgataatcttattatctgggcatttttcaacgcgaaaaacaag

MAD7 - Inscripta   1681 ccggacaagaagattatcgagggtaatacgtcagaaaataagggtgactacaaaaagatgatttataatttgctcccggg
MAD7 - B. subtil   1681 ccggacaaaaaaatcatcgaaggcaatacgtcagagaacaaaggcgactataaaaagatgatctataatctgcttccggg
dMAD7              1681 ccggacaaaaaaatcatcgaaggcaatacgtcagagaacaaaggcgactataaaaagatgatctataatctgcttccggg

MAD7 - Inscripta   1761 tcccaacaaaatgatcccgaaagttttcttgagcagcaagacgggggtggaaacgtataaaccgagcgcctatatcctag
MAD7 - B. subtil   1761 accgaataaaatgatcccgaaagttttctgtcaagcaaaacaggcgtcgaaacatataaaccgtcagcgtatattctgg
dMAD7              1761 accgaataaaatgatcccgaaagttttctgtcaagcaaaacaggcgtcgaaacatataaaccgtcagcgtatattctgg

MAD7 - Inscripta   1841 aggggtataaacagaataaacatatcaagtcttcaaaagactttgatatcactttctgtcatgatctgatcgactacttc
MAD7 - B. subtil   1841 aaggctacaaacagaacaaacacatcaaaagcagcaaggactttgacatcacattttgccatgatctgatcgactacttt
dMAD7              1841 aaggctacaaacagaacaaacacatcaaaagcagcaaggactttgacatcacattttgccatgatctgatcgactacttt

MAD7 - Inscripta   1921 aaaaactgtattgcaattcatcccgagtggaaaaacttcggttttgattttagcgacaccagtacttatgaagacatttc
MAD7 - B. subtil   1921 aagaactgtatcgcaattcatccggaatggaaaaacttcggctttgattttcagacacgagcacgtatgaagatatcag
dMAD7              1921 aagaactgtatcgcaattcatccggaatggaaaaacttcggctttgattttcagacacgagcacgtatgaagatatcag

MAD7 - Inscripta   2001 cgggttttatcgtgaggtagagttacaaggttacaagattgattggacatacattagcgaaaaagacattgatctgctgc
MAD7 - B. subtil   2001 cggctttatagagaagttgaactgcagggctataaaatcgactggacatatatcagcgaaaaggatattgatctgctgc
dMAD7              2001 cggctttatagagaagttgaactgcagggctataaaatcgactggacatatatcagcgaaaaggatattgatctgctgc

MAD7 - Inscripta   2081 aggaaaaaggtcaactgtatctgttccagatatataacaaagatttttcgaaaaaatcaaccgggaatgacaaccttcac
MAD7 - B. subtil   2081 aagaaaaaggccaactgtacctgtttcagatctacaacaaagacttcagcaaaaaaagcacgggcaatgataacctgcat
dMAD7              2081 aagaaaaaggccaactgtacctgtttcagatctacaacaaagacttcagcaaaaaaagcacgggcaatgataacctgcat

MAD7 - Inscripta   2161 accatgtacctgaaaaatcttttctcagaagaaaatcttaaggatatcgtcctgaaacttaacggcgaagcggaaatctt
MAD7 - B. subtil   2161 acgatgtacctgaaaaacctttttagcgaagagaacctgaaagacattgtcctgaaactgaatggcgaagccgaaatttt
dMAD7              2161 acgatgtacctgaaaaacctttttagcgaagagaacctgaaagacattgtcctgaaactgaatggcgaagccgaaatttt

MAD7 - Inscripta   2241 cttcaggaagagcagcataaagaacccaatcattcataaaaaaggctcgattttagtcaaccgtacctacgaagcagaag
MAD7 - B. subtil   2241 ctttcgcaaatccagcattaaaaacccgatcatccataaaaaaggcagcattctggttaaccgcacatatgaagcggaag
dMAD7              2241 ctttcgcaaatccagcattaaaaacccgatcatccataaaaaaggcagcattctggttaaccgcacatatgaagcggaag

MAD7 - Inscripta   2321 aaaaagaccagtttggcaacattcaaattgtgcgtaaaaatattccggaaaacatttatcaggagctgtacaaatacttc
MAD7 - B. subtil   2321 aaaaagatcagtttggcaacattcagatcgtccgcaaaaacattccggaaaacatttatcaagagctgtataaatacttt
dMAD7              2321 aaaaagatcagtttggcaacattcagatcgtccgcaaaaacattccggaaaacatttatcaagagctgtataaatacttt

MAD7 - Inscripta   2401 aacgataaaagcgacaaagagctgtctgatgaagcagccaaactgaagaatgtagtgggacaccacgaggcagcgacgaa
MAD7 - B. subtil   2401 aacgataaaagcgataaagaactgtccgacgaagcagcgaaacttaaaaatgttgttggccatcatgaagcggcaacaaa
dMAD7              2401 aacgataaaagcgataaagaactgtccgacgaagcagcgaaacttaaaaatgttgttggccatcatgaagcggcaacaaa

MAD7 - Inscripta   2481 tatagtcaaggactatcgctacacgtatgataaatacttccttcatatgcctattacgatcaatttcaaagccaataaaa
MAD7 - B. subtil   2481 cattgttaaagactatcgctatacgtacgacaagtacttttctgcatatgccgatcacgattaacttcaaagcaaataaaa
dMAD7              2481 cattgttaaagactatcgctatacgtacgacaagtacttttctgcatatgccgatcacgattaacttcaaagcaaataaaa

MAD7 - Inscripta   2561 cgggttttattaatgataggatcttacagtatatcgctaaagaaaaagacttacatgtgatcggcattgatcggggcgag
MAD7 - B. subtil   2561 cgggctttatcaacgatcgcattctgcagtatattgccaaagaaaaggatctgcatgtcatcggcattgatagaggcgaa
dMAD7              2561 cgggctttatcaacgatcgcattctgcagtatattgccaaagaaaaggatctgcatgtcatcggcattgctagaggcgaa

MAD7 - Inscripta   2641 cgtaacctgatctacgtgtccgtgattgatacttgtggtaatatagttgaacagaaaagctttaacattgtaaacggcta
MAD7 - B. subtil   2641 cgcaatctgatttatgtcagcgttattgatacatgcggcaacattgtcgaacagaaaagctttaacattgtcaacggcta
dMAD7              2641 cgcaatctgatttatgtcagcgttattgatacatgcggcaacattgtcgaacagaaaagctttaacattgtcaacggcta

MAD7 - Inscripta   2721 cgactatcagataaaactgaaacaacaggagggcgctagacagattgcgcggaaagaatggaaagaaattggtaaaatta
MAD7 - B. subtil   2721 tgactaccagatcaagctgaaacagcaagaaggcgcaagacaaattgctcgcaaagaatggaaagaaatcggcaagatca
dMAD7              2721 tgactaccagatcaagctgaaacagcaagaaggcgcaagacaaattgctcgcaaagaatggaaagaaatcggcaagatca

MAD7 - Inscripta   2801 aagagatcaaagagggctacctgagcttagtaatccacgagatctctaaaatggtaatcaaatacaatgcaattatagcg
MAD7 - B. subtil   2801 aagaaattaaagagggctatctgagcctggtcattcatgaaatttctaaaatggtcatcaaatataacgcgattatcgcc
dMAD7              2801 aagaaattaaagagggctatctgagcctggtcattcatgaaatttctaaaatggtcatcaaatataacgcgattatcgcc

MAD7 - Inscripta   2881 atggaggatttgtcttatggttttaaaaaagggcgctttaaggtcgaacggcaagtttaccagaaatttgaaaccatgct
MAD7 - B. subtil   2881 atggaagatctgtcatatggctttaagaaaggccgttttaaagtcgaaagacaggtctaccagaaatttcgaaacaatgct
dMAD7              2881 atggcagatctgtcatatggctttaagaaaggccgttttaaagtcgaaagacaggtctaccagaaatttcgaaacaatgct

MAD7 - Inscripta   2961 catcaataaactcaactatctggtatttaaagatatttcgattaccgagaatggcggtctcctgaaaggttatcagctga
MAD7 - B. subtil   2961 gattaacaaactgaattatctggtgtttaaagacatcagcatcacggaaaatggcggactgctgaaaggctatcaactga
dMAD7              2961 gattaacaaactgaattatctggtgtttaaagacatcagcatcacggaaaatggcggactgctgaaaggctatcaactga

MAD7 - Inscripta   3041 catacattcctgataaacttaaaaacgtgggtcatcagtgcggctgcatttttatgtgcctgctgcatacacgagcaaa
MAD7 - B. subtil   3041 catatattccggataagcttaaaaacgtcggccatcaatgcggctgcatctttttatgttccggcagcgtatacatcaaaa
dMAD7              3041 catatattccggataagcttaaaaacgtcggccatcaatgcggctgcatctttttatgttccggcagcgtatacatcaaaa
```

```
MAD7 - Inscripta   3121 attgatccgaccaccggctttgtgaatatctttaaatttaaagacctgacagtggacgcaaaacgtgaattcattaaaaa
MAD7 - B. subtil   3121 attgatccgacaacaggctttgtcaacatcttcaaattcaaagatctgacggtcgatgcgaaacgcgaattcattaagaa
dMAD7              3121 attgatccgacaacaggctttgtcaacatcttcaaattcaaagatctgacggtcgatgcgaaacgcgaattcattaagaa

MAD7 - Inscripta   3201 atttgactcaattcgttatgacagtgaaaaaaatctgttctgctttacatttgactacaataacttttattacgcaaaaca
MAD7 - B. subtil   3201 atttgacagcatccgctacgacagcgagaaaaatcttttctgctttacgttcgactacaacaacttttatcacgcagaata
dMAD7              3201 atttgacagcatccgctacgacagcgagaaaaatcttttctgctttacgttcgactacaacaacttttatcacgcagaata

MAD7 - Inscripta   3281 cggtcatgagcaaatcatcgtggagtgtgtatacatacggcgtgcgcatcaaacgtcgctttgtgaacggccgcttctca
MAD7 - B. subtil   3281 cggttatgtcaaaaagcagctggtcagtctatacatatggcgttagaattaaacgcagatttgtgaacggcagatttagc
dMAD7              3281 cggttatgtcaaaaagcagctggtcagtctatacatatggcgttagaattaaacgcagatttgtgaacggcagatttagc

MAD7 - Inscripta   3361 aacgaaagtgataccattgacataaccaaagatatggagaaaacgttggaaatgacggacattaactggcgcgatggcca
MAD7 - B. subtil   3361 aatgaaagcgatacaatcgacatcacgaaagacatggaaaaaacgcttgaaatgacggatattaactggcgtgatggaca
dMAD7              3361 aatgaaagcgatacaatcgacatcacgaaagacatggaaaaaacgcttgaaatgacggatattaactggcgtgatggaca

MAD7 - Inscripta   3441 cgatcttcgtcaagacattatagattatgaaattgttcagcacatattcgaaattttccgtttaacagtgcaaatgcgta
MAD7 - B. subtil   3441 tgatcttcgccaggatattatcgattatgaaatcgtccagcacatctttgaaatctttagactgacagtccaaatgcgca
dMAD7              3441 tgatcttcgccaggatattatcgattatgaaatcgtccagcacatctttgaaatctttagactgacagtccaaatgcgca

MAD7 - Inscripta   3521 actccttgtctgaactggaggaccgtgattacgatcgtctcatttcacctgtactgaacgaaaataacatttttatgac
MAD7 - B. subtil   3521 attcactgtcagaacttgaagatagagattatgatcgcctgatttctccggtcctgaatgaaaataacatctttacgat
dMAD7              3521 attcactgtcagaacttgaagatagagattatgatcgcctgatttctccggtcctgaatgaaaataacatctttacgat

MAD7 - Inscripta   3601 agcgcgaaagcgggggatgcacttcctaaggatgccgatgcaaatggtgcgtattgtattgcattaaaagggttatatga
MAD7 - B. subtil   3601 agcgcaaaagcaggcgacgcactgccgaaagatgcggatgcaaatggcgcatattgcattgcactgaaaggcctgtatga
dMAD7              3601 agcgcaaaagcaggcgacgcactgccgaaagatgcggctgcaaatggcgcatattgcattgcactgaaaggcctgtatga

MAD7 - Inscripta   3681 aattaaacaaattaccgaaaattggaaagaagatggtaaatttttcgcgcgataaactcaaaatcagcaataaagattggt
MAD7 - B. subtil   3681 aatcaaacaaatcacagaaaactggaaagaggacggcaaatttttctcgcgataaactgaaaatcagcaacaaggactggt
dMAD7              3681 aatcaaacaaatcacagaaaactggaaagaggacggcaaatttttctcgcgataaactgaaaatcagcaacaaggactggt

MAD7 - Inscripta   3761 tcgactttatccagaataagcgctatctctaa
MAD7 - B. subtil   3761 ttgacttcatccaaaataagcgctacctgtaa
dMAD7              3761 ttgacttcatccaaaataagcgctacctgtaa
```

Appendix E

Copy of “CRISPR-Cas9 *In Situ* engineering of subtilisin E in *Bacillus subtilis*” by Marcus A. Price, Rita Cruz, Scott Baxter, Franck Escalettes and Susan J. Rosser. Published on 7th January 2019 in PLoS ONE.

PLOS ONE

RESEARCH ARTICLE

CRISPR-Cas9 *In Situ* engineering of subtilisin E in *Bacillus subtilis*

Marcus A. Price[1,2]*, **Rita Cruz**[2], **Scott Baxter**[2], **Franck Escalettes**[2☯], **Susan J. Rosser**[1☯]*

1 Department of Quantitative Biology, Biochemistry and Biotechnology, University of Edinburgh, Edinburgh, United Kingdom, **2** Ingenza Ltd., Roslin Innovation Centre, Roslin, United Kingdom

☯ These authors contributed equally to this work.
* M.A.Price-4@sms.ed.ac.uk (MAP); Susan.Rosser@ed.ac.uk (SJR)

Citation: Price MA, Cruz R, Baxter S, Escalettes F, Rosser SJ (2019) CRISPR-Cas9 *In Situ* engineering of subtilisin E in *Bacillus subtilis*. PLoS ONE 14(1): e0210121. https://doi.org/10.1371/journal.pone.0210121

Editor: Alfred S. Lewin, University of Florida, UNITED STATES

Received: September 21, 2018

Accepted: December 17, 2018

Published: January 7, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by the Industrial Biotechnology Innovation Centre (IBioIC: www.ibioic.com) and Ingenza limited (www.ingenza.com) as part of M.A.P PhD studentship. Ingenza Ltd. provided support in the form of salaries for authors (R.C., S.B. and F.E.), but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of

Abstract

CRISPR-Cas systems have become widely used across all fields of biology as a genome engineering tool. With its recent demonstration in the Gram positive industrial workhorse *Bacillus subtilis*, this tool has become an attractive option for rapid, markerless strain engineering of industrial production hosts. Previously described strategies for CRISPR-Cas9 genome editing in *B. subtilis* have involved chromosomal integrations of Cas9 and single guide RNA expression cassettes, or construction of large plasmids for simultaneous transformation of both single guide RNA and donor DNA. Here we use a flexible, co-transformation approach where the single guide RNA is inserted in a plasmid for Cas9 co-expression, and the donor DNA is supplied as a linear PCR product observing an editing efficiency of 76%. This allowed multiple, rapid rounds of *in situ* editing of the subtilisin E gene to incorporate a salt bridge triad present in the *Bacillus clausii* thermotolerant homolog, M-protease. A novel subtilisin E variant was obtained with increased thermotolerance and activity.

Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) systems of adaptive immunity in bacteria have become widely used across all fields of biology as a genome editing tool since demonstration of its use as a RNA-guided DNA endonuclease in 2012 [1]. The technique, based on the type II CRISPR-Cas9 system from *Streptococcus pyogenes*, makes use of the host DNA double-strand break (DSB) repair machinery to introduce mutations within the DNA sequence. DSBs can be repaired by the error-prone non-homologous end joining (NHEJ) or homology directed DNA repair (HDR) mechanisms of the cell. NHEJ can introduce random mutations during repair, while HDR allows the introduction of selected modifications in the presence of an engineered template with chromosome homology regions flanking the desired mutations. Uptake of CRISPR-Cas9 genome editing in the Gram positive model organism *Bacillus subtilis* was however relatively slow, perhaps, in part, due to the high number of well-established genomic modification tools already available. However, recently several publications were released in quick succession showing the development of the CRISPR-Cas9 system in this host [2–6]. These initially showed use of the catalytically active Cas9 for large scale genomic deletions, small and large DNA insertions, gene silencing by

these authors are articulated in the 'author contributions' section.

Competing interests: The authors declare R.C., S. B. and F.E.'s affiliation to Ingenza Ltd. This affiliation does not alter our adherence to PLOS ONE policies on sharing data and materials. The authors also declare that no competing interests exist, there are no patents pending for this research, no products in development of marketed.

introduction of a stop codon and engineering of increased resistance to bacteriophage SPP1 [2–6]. Additionally, the catalytically inactivated 'dead' Cas9 (dCas9) was used to allow phenotypic analysis of all essential genes in *B. subtilis*. This was performed by dCas9 binding at the target region within the gene and not inducing a DSB, subsequently reducing the level of transcription from that gene and any other genes downstream in the same operon. [7].

Two methodologies have been exemplified in *B. subtilis* using the catalytically active Cas9. First, Westbrook *et al.* (2016) demonstrated a system where a single plasmid simultaneously delivers an integration cassette for expression of the chromosome targeting crRNA and the donor DNA (dDNA) for DSB repair, in a strain where the *cas9* gene and tracrRNA, which associates Cas9 to the crRNA, was previously integrated in the chromosome [3]. Others have used an alternative system where *cas9* and sgRNA (single guide RNA where tracrRNA and crRNA are combined as a single fragment) expression cassettes, as well as dDNA, are delivered in a single temperature sensitive plasmid [2,5,6]. With the first method, the crRNA expression cassette must be removed with a subsequent transformation step to restore the integration site before next rounds of editing are possible. The second strategy, although more amenable to multiple rounds of editing, involves a 2-step construction of a large shuttle vector to introduce the sgRNA and dDNA [2]. Here, we present a simple and straightforward method which allows rapid *in situ* edits of the genome, using a plasmid which can target any loci within *B. subtilis* in a single cloning step, and without previous genome modifications. During the editing process, this plasmid transcribes the sgRNA under the control of the strong constitutive promoter, P_{veg}, and utilises the basal expression level of the P_{grac} promoter to regulate *cas9* expression. The dDNA in the form of a linear PCR product is co-transformed with the plasmid, providing the template for DSB repair.

The use of CRISPR-Cas9 mediated genome editing using this approach was first exemplified by knock-out of the α-amylase encoding gene, *amyE*. Subsequently our system was shown to be compatible for engineering of industrial production hosts by *in situ* modification of the *aprE* gene. This encodes for the industrially relevant enzyme subtilisin E, used globally within the detergent industry. As the wild type enzyme can be broken down by detergent formulations and heat, subtilisin variants with improved thermostability and pH tolerance have long been of interest. Subtilisin E has been widely used as a target for protein engineering experiments [8]. A salt-bridge triad (Arg19-Glu271-Arg275), identified in the subtilisin E homolog from *Bacillus clausii* (M-protease), was found to be a contributor for the characteristic thermotolerance of this enzyme [9]. Here, we use CRISPR-Cas9 mediated genome editing to replace the respective residues in subtilisin E (Gln125-Gln377-Gln381), and evaluate the effect of the salt bridge on the thermostability and activity of the new variant. This work illustrates the use of a simple CRISPR-Cas9 system for *B. subtilis* rapid, *in situ* protein engineering, which is at the core of industrial biotechnology to provide new, suitable and competitive biocatalysts.

Materials and methods

Strains and media

The strains and plasmids used in this study are listed in Table 1. All oligonucleotides used in this study are listed in S1 Table. *Escherichia coli* Top10 strains were used to construct recombinant plasmids. Bacterial cells were cultured in Lysogeny broth (LB) or LB agar (VWR) media at 37˚C. Unless otherwise stated, the following antibiotics were added to the media when required: ampicillin (200 μg/mL), chloramphenicol (10 μg/mL), or spectinomycin (100 μg/mL).

Plasmid construction

Unless otherwise stated, plasmid construction was performed as described in Sambrook and Russell [12]. DNA oligonucleotides were purchased from Sigma-Aldrich. The reagents for

Table 1. Strains and plasmids used in this study.

Strain/ Plasmid	Genotype	Reference
Strains		
B. subtilis		
168	*trpC2*	Laboratory stock
BAC0094	*B. subtilis* 168 with *aprE* Q125R variant.	This work
BAC0095	*B. subtilis* 168 with *aprE* Q377E and Q381R variant.	This work
BAC0097	BAC0095 with *aprE* Q125R, Q377E and Q381R variant	This work
BAC0114	*B. subtilis* 168 *ΔaprE::aad9*	This work
BAC0119	BAC0114, pBAC0059.	This work
BAC0120	BAC0114, pBAC0060.	This work
BAC0121	BAC0114, pBAC0068.	This work
BAC0122	BAC0114, pBAC0069.	This work
E. coli		
Top10	*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu) 7697 galU galK rpsL* (Str^R) *endA1 nupG*	Invitrogen
Plasmids		
pHT01	*E. coli/B. subtilis* shuttle vector carrying P_{grac} and *lacI*, *bla*, *cat*.	MoBiTec
pdCas9-bacteria	*tetR*; *dCas9* (*S. pyogenes*); rrnB T1; p15a ori; *cat*	[10]
pING0001	P_{bad}; rrnB T1; rrnB T2; *bla;* pBR327 ori; *araC*	Ingenza Ltd.
pING0002	P_{bad}; *aphAI;* pMB1 ori; *araC*	Ingenza Ltd.
pDR111	*bla*; 5' *amyE*; *aad9*; P_{spac};*lacI*; 3' *amyE*	[11]
pBAC0001	pHT01 with SapI sites removed.	This work
pBAC0008	pBAC0001 with rrnB T1 T2; sgRNA Cas9 handle; protospacer cloning site; P_{veg}	This work
pBAC0013	pdCas9-bacteria with catalytically active *cas9*.	This work
pBAC0015	pBAC0008 with *cas9* (*S. pyogenes*) from pBAC0013. Cas9 expression regulated by the P_{grac} promoter.	This work
pBAC0027	pBAC0015 with amyE-1 sgRNA DNA.	This work
pBAC0035	pBAC0015 with non-targeting sgRNA DNA.	This work
pBAC0041	pBAC0015 with amyE-2 sgRNA DNA.	This work
pBAC0047	pBAC0015 with amyE-3 sgRNA DNA.	This work
pBAC0054	pBAC0015 with aprE-1 sgRNA DNA.	This work
pBAC0055	pBAC0015 with aprE-2 sgRNA DNA.	This work
pBAC0059	pHT01 with *aprE* (native) from *B. subtilis* 168 with ATG start codon.	This work
pBAC0060	pHT01 with *aprE* (Q125R, Q377E and Q381R variant) from BAC0097 with ATG start codon.	This work
pBAC0068	pHT01 with *aprE* (Q125R variant) from BAC0094 with ATG start codon.	This work
pBAC0069	pHT01 with *aprE* (Q377E and Q381R variant) from BAC0095 with ATG start codon.	This work

https://doi.org/10.1371/journal.pone.0210121.t001

PCR, restriction digest, Gibson assembly, DNA phosphorylation and ligation were purchased from New England Biolabs (NEB). DNA purification was performed using the kits from NEB. DNA sequences were confirmed by Source Bioscience.

pBAC0001 was constructed by Gibson assembly to remove the SapI sites present in pHT01 (MoBiTec) [13]. PCR products for Gibson assembly were prepared from pHT01 using oligonucleotides oMAP0002/0003/0004/0005.

pBAC0008 plasmid was constructed using the inABLE plasmid assembly method [14]. Individual 5' truncated parts were prepared by PCR followed by restriction digest at 5' and 3' regions with SapI. These parts were ligated to phosphorylated and annealed oligonucleotides at each terminus, containing 3 nt and 16 nt single strands at the 5' and 3' ends, respectively. The part- oligonucleotides fusions were annealed at the homologous 16 nt overhangs for 1 hour at 37˚C, and used to transform electrocompetent *E. coli*.

pBAC0008 consisted of four parts: 1. the *E. coli/B. subtilis* shuttle vector backbone from pBAC0001; 2. the LacI repressor and isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible P_{grac} promoter from pBAC0001, including a multiple cloning site; 3. the bidirectional strong *rrnB* T1 and T2 terminators; 4. the sgRNA expression module consisting of a kanamycin resistance gene flanked by AarI sites expressed under the control of the P_{veg} promoter, and the 'Cas9 handle' section of the sgRNA (5' truncated parts were prepared by PCR from the indicated template and oligonucleotides: 1. pBAC0001 with oMAP0010/0011; 2. pBAC0001 with oMAP0018/0019; 3. pING0001 with oMAP0024/0025; 4. pING0002 with oMAP0030/0031. Parts were ligated at 5' and 3' respectively with annealed oligonucleotides: 1. oMAP0008/0009 and oMAP0014/0015; 2. oMAP0016/0017 and oMAP0020/0021; 3. oMAP0022/0023 and oMAP0048/0049; 4. oMAP0050/0051 and oMAP0052/0053.)

pBAC0013 was constructed by converting the catalytically inactive *dcas9* gene from pdCas9-bacteria to active *cas9*. This was done by introducing mutations A10D, A840H and removing a BamHI site by PCR (oligonucleotides oMAP0062/0063/0064/0065/0066/0067) and a subsequent 3-part Gibson Assembly (pdCas9-bacteria was a gift from Stanley Qi (Addgene plasmid # 44249) [10].

Subsequently the catalytically active *cas9* gene was amplified with oligonucleotides oMAP0073/0074, introducing a BsaI site and XbaI recognition sites at the 5' and 3' end of the gene, respectively. The amplified *cas9* was digested with BsaI and XbaI, pBAC0008 was digested with BamHI and XbaI and both fragments were ligated with T4 DNA ligase, yielding pBAC0015.

To construct the subtilisin E overexpression plasmids pBAC0059, pBAC0060, pBAC0068 and pBAC0069, each subtilisin E variant and the wild type (WT) gene was amplified from chromosomal DNA derived from the relevant strain (Table 1) with oligonucleotides oMAP0186/0203, oMAP0200/0203, oMAP0186/0203, and oMAP0200/0203 respectively, each set containing 5'-BamHI and 3'-XmaI recognition sites. Oligonucleotides oMAP0203 also replaced the native GTG start codon with an ATG codon. The genes were cloned into pHT01 using the BamHI and XmaI recognition sites in the multiple cloning site of the plasmid.

CRISPR-Cas9-Mediated gene editing in *B. subtilis*

Analysis of editing efficiency. For CRISPR-Cas9 mediated editing efficiency analysis, plasmids targeting the α-amylase *amyE* in *B. subtilis* were constructed. Plasmid pBAC0015 was digested with AarI (Thermo Fisher Scientific) to remove the kanamycin resistance gene and ligated to two annealed and phosphorylated DNA oligonucleotides of 24 nucleotides. The annealed oligonucleotides consisted of the 20 nucleotide protospacer region for targeting of the Cas9 protein and 4 nucleotides to generate a single stranded overhang for compatibility with the AarI digested vector. DNA phosphorylation was performed using T4 Polynucleotide Kinase according to the manufacturer's instructions. The protospacer regions were identified using the online tool CRISPR-ERA: a comprehensive designer tool for CRISPR genome editing, (gene) repression, and activation, and selected based on the proximity to the desired modification [15]. The oligonucleotide pairs oMAP0089/0091, oMAP0125/0127 and oMAP0140/0142 yielded plasmids pBAC0027, pBAC0041 and pBAC0047 respectively. As a positive control for transformation efficiency, a sgRNA designed not to target the *B. subtilis* 168 chromosome was inserted into pBAC0015 with the oligonucleotide pair oMAP0145/0147, yielding pBAC0035. dDNA to introduce stop codons and repair the sgRNA-targeted Cas9 DSB were constructed by overlap extension PCR (OE-PCR) of two DNA fragments as described by Bryksin and Matsumura [16]. The desired edits were introduced in the homology overlap of the two DNA fragments. These included the silent mutation of the 5'-NGG-3' protospacer

adjacent motif (PAM) site to prevent continuous cutting by the RNA-guided Cas9 endonuclease. OE-PCR products from oligonucleotides sets oMAP0121/0122/0123/0124, oMAP0121/0122/0128/0129 and oMAP0121/0122/0143/0144 were used alongside pBAC0027, pBAC0041 and pBAC0047 respectively. Genome editing was carried out in triplicate by co-transforming naturally competent *B. subtilis* 168 with 200 ng CRISPR-Cas9/sgRNA plasmid DNA and 1 μg dDNA [17]. Transformants were spread on LB agar plates supplemented with 1% soluble potato starch (VWR) and chloramphenicol (5 μg/mL). IPTG was not added for *cas9* expression induction to limit cellular burden and it was found that basal expression under the P_{grac} promoter was sufficient for *cas9* expression. Effective knock-out of *amyE* by stop codon introduction was determined by staining transformation plates with iodine [18].

Editing of *aprE*. Using the approach described above, CRISPR-Cas9 mediated editing of the *aprE* gene in *B. subtilis* 168 was carried out. The oligonucleotide pairs oMAP0150/0151 and oMAP0156/0157 were ligated with pBAC0015 prepared as above, yielding plasmids pBAC0054 and pBAC0055 respectively. dDNA OE-PCR products from oligonucleotides sets oMAP0152/0153/0154/0155 and oMAP0158/0159/0160/0161 were used alongside pBAC0054 and pBAC0055 respectively. Editing was confirmed by PCR of the *aprE* gene with oligonucleotides oMAP0152/0161, followed by sequencing using oligonucleotides oMAP0158 and oMAP0155 for candidates edited with pMAP0054 and pMAP0055 respectively. The CRISPR-Cas9/sgRNA plasmid was removed from the edited strain by promoting plasmid loss in LB supplemented with 1 mM IPTG overnight. Here the presence of IPTG encourages curing by introducing the pressure of protein expression without the presence of chloramphenicol to retain the plasmid within the cell. Plasmid loss was confirmed by counter plating on LB agar plates with and without chloramphenicol before a subsequent round of editing. Following the second round of editing, these mutations and those introduced previously were confirmed to be present by sequencing (as above).

Protein purification

The three subtilisin E variants and WT protein were overexpressed in the *aprE* knock-out strain BAC0114. BAC0114 was constructed by transformation of *B. subtilis* 168 with OE-PCR product containing a spectinomycin resistance cassette flanked by homology arms upstream and downstream of *aprE* (oligonucleotides oMAP0217/0218/0219/0220/0221/0222). Confirmation of *aprE* deletion was obtained by purification of the genomic DNA for BAC0114 and PCR of *ΔaprE*::*aad9* locus with oligonucleotides oMAP0670 (hybridising to *aad9*) and oMAP0671 (hybridising to the genome, upstream of the homology arm region). Additionally, PCR with oligonucleotide pair oMAP0217/0220 (hybridising to the extremities of the homology arm region) revealed the expected increase in product size for *ΔaprE*::*aad9* relative to the WT PCR product. This strain was then transformed with plasmids pBAC0059/0060/0068/0069, resulting in strains BAC0119/0120/0121/0122 respectively. These strains were grown for 24 hours at 37˚C with agitation (250 rpm) in 20 mL LB supplemented with chloramphenicol. The supernatant was clarified and dialysed overnight into 100 mM Tris-HCl with 150 mM sodium chloride at pH 8. The dialysed supernatant was concentrated to 5 mL with Amicon Ultra-15 (10 kDa membrane (Merck)), filtered through a 0.2 μm filter (Sartorius). This was loaded onto a HiPrep Sephacryl S-200 HR, 120 mL 16/60 size exclusion column (GE Healthcare), eluted in the same buffer composition used for dialysis and fractions found to contain subtilisin E were pooled. Protein concentration was determined by absorbance at 280 nm.

Thermal shift assay

The fluorescence-based thermal shift assay was used to determine the melting temperature (T_m) of the subtilisin E variants and WT [19]. The assay reaction was prepared in a total

volume of 20 μL (18 μL purified subtilisin E variant in purification buffer, 1 μL 100 mM calcium acetate and 1 μL 1:50 SYPRO Orange (Sigma-Aldrich)). The assay was performed with six technical replicates for each variant. The reaction was analysed using the Pikoreal 96 (Thermo Scientific) which recorded changes in fluorescence with increasing temperature (40˚C—80˚C in increments of 0.2˚C, held for 6 seconds at each point).

Subtilisin E activity assay

Subtilisin E variants and WT activity was determined by casein degradation as described by Cupp-Enyard [20].

Residual enzyme activity tests

Residual enzyme activity was analysed by incubation of the purified enzymes at 55˚C for 10, 20, 40, and 60 minutes prior to the activity assay.

Results

Analysis of CRISPR-Cas9 mediated gene editing efficiency

To analyse CRISPR-Cas9 mediated gene editing efficiency using our co-transformation system, the *amyE* gene encoding an α-amylase was selected for knock-out using a starch degradation assay. The three tested *amyE* targeting plasmids (pBAC0027/0041/0047) yielded on average an editing efficiency of 63.9%, 89.2% and 74.4% when co-transformed with dDNA (Fig 1).

Rapid *in situ* modification of *aprE* with CRISPR-Cas9

The crystal structures for *B*. *subtilis* subtilisin E (PDB ID 1SCJ) and its *B*. *clausii* homolog, M-protease (PDB ID 1WSD), were overlaid using Swiss-Pdb viewer (Fig 2) [21–23]. The residues corresponding to the salt-bridge triad (R19-E271-R275) that have previously been shown to contribute towards M-protease thermotolerance were identified in the subtilisin E structure (Q125-Q377-Q381) [9].

A CRISPR-Cas9 mediated chromosome editing strategy was designed to introduce the mutations Q125R, Q377E and Q381R. pBAC0015 was altered to include the oligonucleotide pairs oMAP0150/oMAP0151 or oMAP0156/oMAP0157, yielding plasmids pBAC0054 (targeting Q125R locus) and pBAC0055 (targeting Q377E and Q381R locus) respectively. Due to their proximity, Q377E and Q381R modifications were combined into a single CRISPR-Cas9 mediated editing step with a single dDNA. Q125R was modified in a separate editing step. The desired edits, included the modification of the PAM site to prevent continuous cutting by the RNA-guided Cas9 endonuclease, were introduced in the homology overlap of the two DNA fragments prepared using PCR. These were subsequently combined using OE-PCR and co-transformed alongside the respective plasmid (Fig 2). OE-PCR products from oligonucleotides sets oMAP0152/0153/0154/0155 and oMAP0158/0159/0160/0161 were used alongside pBAC0054 and pBAC0055 respectively. Following an efficient curing process yielding strains BAC0094 (Q125R) and BAC0095 (Q377E and Q381R) respectively, a second round of editing using the second set of editing plasmid and dDNA yielded the final strain (BAC0097) containing all three modified residues. All screened colonies were found to contain the desired mutations following sequencing.

This system represents a rapid technique for *in situ* protein modifications within *B*. *subtilis* 168. Once the Cas9-sgRNA expression plasmid is prepared, the same region on the

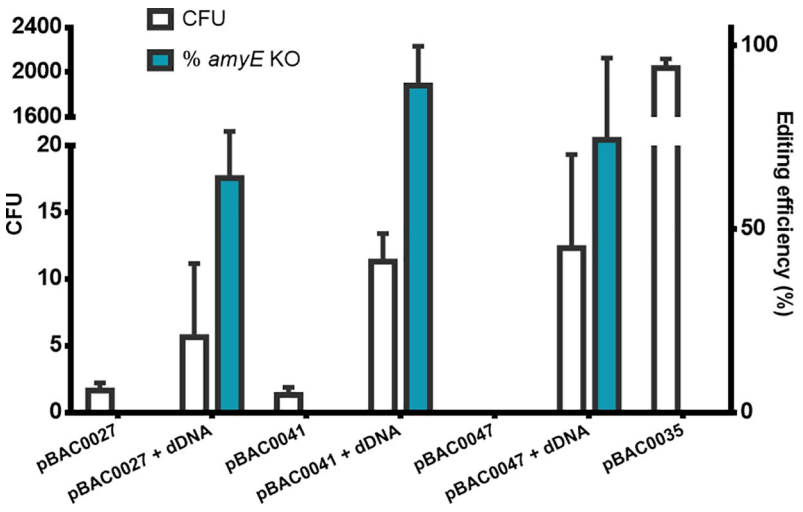


Fig 1. Number of transformants and editing efficiencies obtained following CRISPR-Cas9 genome editing of *amyE*. White bars represent the number of CFU obtained following transformation of three *amyE* targeting plasmids (pBAC0027/0041/0047) with or without editing template (dDNA) to introduce stop codons and repair the sgRNA-targeted Cas9 DSB. A plasmid containing a non-targeting sgRNA (pBAC0035) was transformed to monitor transformation efficiency. The blue bars represent the editing efficiency of the obtained CFU determined by observing the presence or absence of a halo following iodine staining of the starch-containing transformation plates. Error bars indicate the standard deviation between three transformations.

https://doi.org/10.1371/journal.pone.0210121.g001

chromosome can be targeted with alternative, rapidly prepared, linear dDNA templates conveying novel modifications of interest.

Thermotolerance of subtilisin E variants

The thermotolerance of purified subtilisin E variants and WT (purification described in materials and methods section) was analysed by the thermal shift assay (Fig 3). Variants Q377E +Q381R and Q125+Q377E+Q381R showed increased thermotolerance when compared to the WT and Q125R proteins, confirming the importance of residues E377 and R381 for protein stability at higher temperatures. In contrast, the Q125R variant showed no increase in thermotolerance, most likely as there is no mutated residue with which it can form salt bridge.

The average T_m for each variant was established by performing a melt curve analysis of the thermal shift assay data (Fig 3). An increase in T_m of up to 1.4˚C was established and verified as statistically significant by use of an unpaired t test with Welch's correction (Fig 4). The greatest increase in T_m of 1.4˚C was found in the Q377E + Q381R variant (54.8˚C, P value = < 0.0001), while the full salt bridge triad variant showed an increase of 1.2˚C (54.6˚C, P value = < 0.0001). The Q125R variant showed no significant difference in T_m when compared to the WT (53.4˚C and 53.5˚C respectively).

Activity retention of subtilisin E variants

To ensure the introduced mutations had not negatively affected catalytic efficiency of the enzyme, the activity for each variant and the WT was established under neutral conditions (pH 7.5, 37˚C) by measuring the degradation of casein using Folin's reagent (Fig 5) [20]. Both the Q125R and Q125+Q377E+Q381R variants were 60% less active when compared to the WT, suggesting that the Q125R residue is important for catalytic activity. The Q377E+Q381R variant showed an increase in protease activity of 46.5% (P value = < 0.0001).

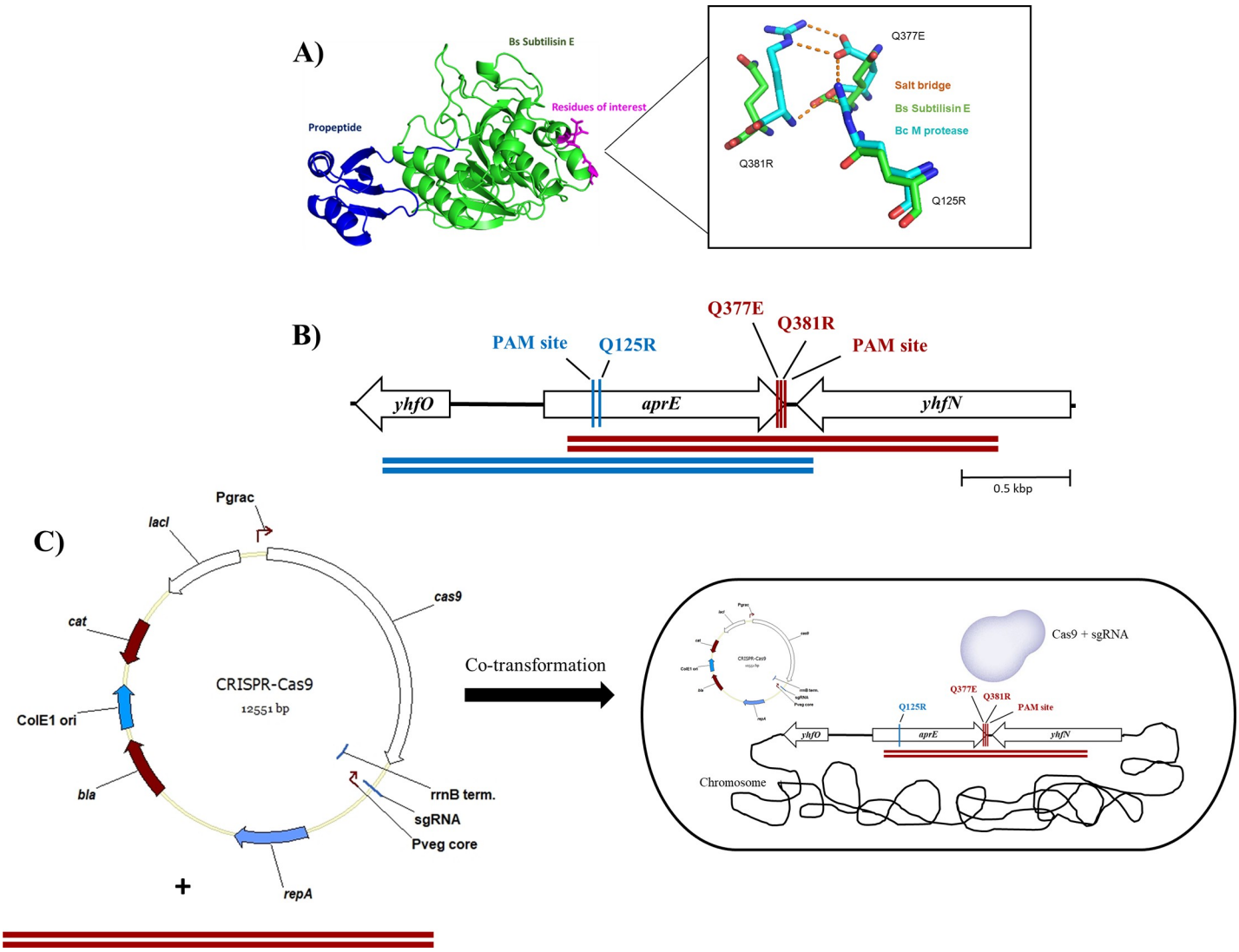


Fig 2. Design and CRISPR-Cas9 editing process of *aprE*. A) *B. subtilis* 168 subtilisin E crystal structure (green) and overlay with *B. clausii* M protease crystal structure (cyan), with the associated salt bridge (dashed yellow line). **B)** Design of the sgRNA and dDNA OE-PCR product for CRISPR-Cas9 genome editing. The two rounds of editing are described in blue and red, the PAM recognition sequences for each sgRNA was also targeted for disruption in each dDNA. **C)** Description of editing process where the second round of editing occurs following a blunt DSB by sgRNA guided Cas9.

https://doi.org/10.1371/journal.pone.0210121.g002

Residual enzyme activity was measured following incubation of the enzyme variants at 55˚C (Fig 6). Variant Q377E + Q381R was found to be the best at retaining its enzymatic activity when incubated for over 20 minutes. An improvement of 12.1% and 15.1% in activity relative to the WT following incubation for 40 and 60 minutes respectively was noted. The Q125 + Q377E + Q381R variant showed a 34.9% decreased in activity relative to the WT following 20 minutes incubation. Similarly, the Q125R variant showed a decreased in activity of 54.1% at 20 minutes incubation.

Discussion

We have developed a *B. subtilis* specific, rapid, *in situ* protein engineering technique based on the ground breaking CRISPR-Cas9 technology. The rapid engineering of established and

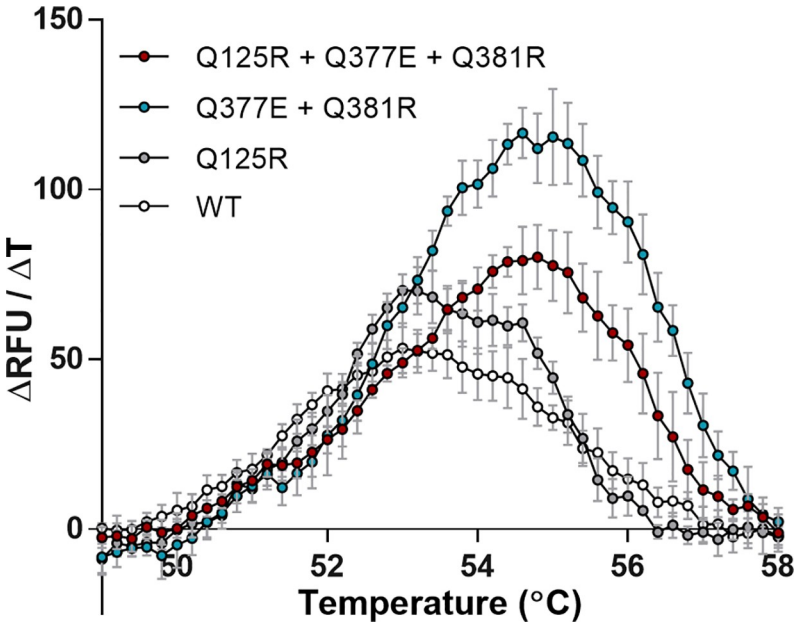


Fig 3. Thermal shift assay of the thermostability of each subtilisin E variant and WT. Increase in fluorescence was detected as hydrophobic regions of the protein were exposed as the protein denatured. The protein melting temperature of each variant and WT was calculated with a melt curve analysis, determining the peak rate of protein unfolding per temperature increase. Error bars indicate the standard deviation between six replicates.

https://doi.org/10.1371/journal.pone.0210121.g003

novel industrially relevant proteins yielding markerless, and potentially non-GMO compliant strains, is an attractive proposition for industrial biotechnology. This is due to the lowering in

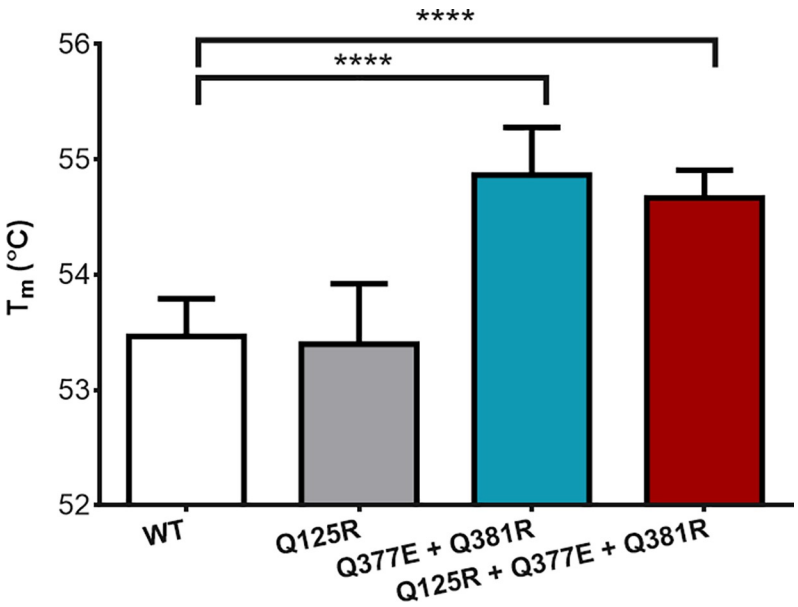


Fig 4. Average T_m between Subtilisin variants and WT replicates. The T_m of each subtilisin E variant and WT was calculated as the temperature at which the peak rate of protein unfolding was observed for each of six replicated. Error bars indicate the standard deviation. **** = p value summary ($p = < 0.0001$) following unpaired t test with Welch's correction.

https://doi.org/10.1371/journal.pone.0210121.g004

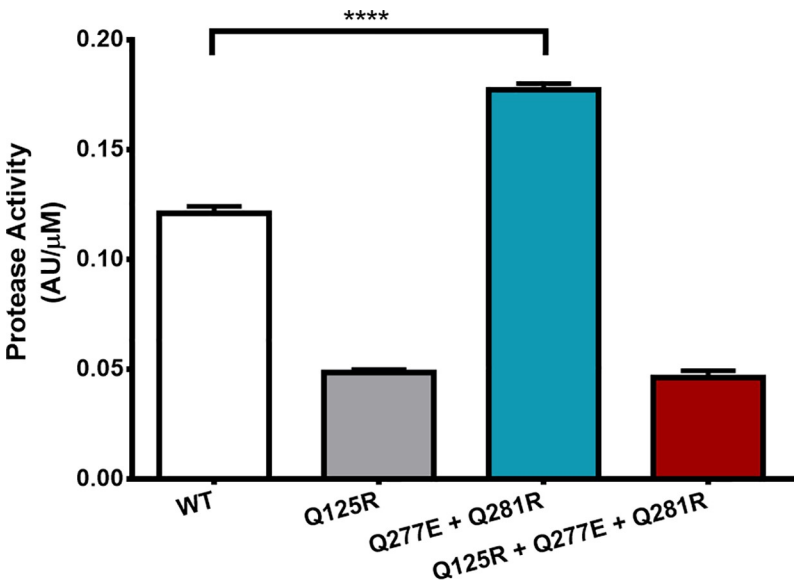


Fig 5. Protease activity assay. Protease activity at pH 7.5, 37˚C for 10 minutes. Normalised for protein concentration determined by absorbance at 280 nm. Absorbance units (AU) determined at 660 nm. **** = p value summary (p = < 0.0001) following unpaired t test with Welch's correction. Error bars indicate standard deviation between triplicates.

https://doi.org/10.1371/journal.pone.0210121.g005

costs and time associated with product development and governmental regulatory compliance, while increasing the time a product is in production.

Initial exemplification of our system on the α-amylase encoding gene, *amyE*, showed an average editing efficiency of 76% across the three targeted PAM sites selected. While our co-transformation-based approach to genome editing may result in lower transformation efficiencies due to the requirement that both plasmid and dDNA molecules enter the cell, the rate of editing efficiency observed enables a high degree of confidence that most colonies obtained

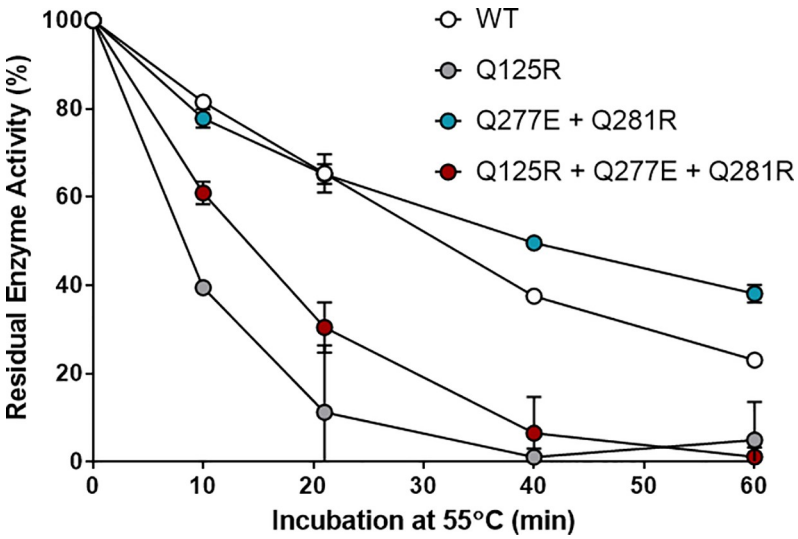


Fig 6. Residual protease activity assay. Residual protease activity under neutral conditions following incubation at 55˚C for various lengths of time, determined by a casein degradation enzymatic assay. Error bars indicate standard deviation between technical triplicates.

https://doi.org/10.1371/journal.pone.0210121.g006

have been edited as designed. This system, where one round of plasmid construction is required and transformed alongside an easily prepared PCR product, represents an alternative to the *B. subtilis* CRISPR-Cas9 genome editing strategies published to date. These involve previous chromosome integration of the *cas9* gene and subsequent eviction of *cas9*/tracrRNA and crRNA expression cassettes after editing, or a two-step plasmid construction to introduce the sgRNA expression cassette and dDNA [2,3,5,6].

In under a week we carried out two rounds of *in situ* CRISPR-Cas9 editing of the subtilisin E gene, *aprE*. A salt bridge triad was introduced and the Q337E + Q381R variant showed an increase of 46.5% in subtilisin E activity, as well as a 1.4˚C increase in thermostability. To the best of our knowledge these modifications have not been combined before in *B. subtilis* subtilisin E.

Interestingly, the Q125R + Q377E + Q381R variant did not retain its activity levels to the same level as the Q377E + Q381R variant. Indeed, in both variants where the Q125R mutation is introduced a drastic drop in enzyme activity retention was observed (Figs 5 and 6), indicating a vital catalytic activity role for residue Q125 within *B. subtilis* subtilisin E. A similar study of salt bridges in subtilisin E was performed by Erwin *et al.*, where mutations Q125E + Q377E formed a salt bridge based on X-ray crystal structure, but resulted in a 1.2˚C drop in thermotolerance [24].

In summary, this method for genomic modifications by CRISPR-Cas9 allows rapid, *in situ* protein engineering of industrially relevant strains. Furthermore, discovery and optimisation of new molecular biology tools such as the work presented here, increases the speed and efficiency at which novel biocatalysts can be developed for sustainable bioprocesses.

Supporting information

S1 Fig. Thermal shift assay of the thermostability of each subtilisin E variant and WT. Fluorescence increase observed as a result of protein unfolding and hydrophobic residue exposure during the thermal shift assay.
(TIF)

S1 File. Thermo denaturation assay and melt curve analysis data.
(XLSX)

S2 File. Subtilisin E protein activity assay data.
(XLSX)

S1 Table. Oligonucleotides used in this study.
(PDF)

Acknowledgments

The authors would like to thank Prof. Colin Harwood for providing plasmid pDR111 for this study.

Author Contributions

Conceptualization: Marcus A. Price, Rita Cruz, Scott Baxter.

Data curation: Marcus A. Price, Rita Cruz.

Formal analysis: Marcus A. Price.

Investigation: Marcus A. Price.

Methodology: Marcus A. Price.

Software: Scott Baxter.

Supervision: Rita Cruz, Scott Baxter, Franck Escalettes, Susan J. Rosser.

Writing – original draft: Marcus A. Price.

Writing – review & editing: Marcus A. Price, Rita Cruz, Scott Baxter, Franck Escalettes, Susan J. Rosser.

References

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided. Science. 2012; 337: 816–822. https://doi.org/10.1126/science.1225829 PMID: 22745249
2. Altenbuchner J. Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system. Appl Environ Microbiol. 2016; 82: 5421–5427. https://doi.org/10.1128/AEM.01453-16 PMID: 27342565
3. Westbrook AW, Moo-Young M, Chou CP. Development of a CRISPR-Cas9 toolkit for comprehensive engineering of *Bacillus subtilis*. Appl Environ Microbiol. 2016; 82(16): 4876–4895. https://doi.org/10.1128/AEM.01159-16 PMID: 27260361
4. Jakutyte-Giraitiene L, Gasiunas G. Design of a CRISPR-Cas system to increase resistance of *Bacillus subtilis* to bacteriophage SPP1. J Ind Microbiol Biotechnol. Springer Berlin Heidelberg; 2016; 43: 1183–1188. https://doi.org/10.1007/s10295-016-1784-0 PMID: 27255973
5. Zhang K, Duan X, Wu J. Multigene disruption in undomesticated *Bacillus subtilis* ATCC 6051a using the CRISPR/Cas9 system. Sci Rep. Nature Publishing Group; 2016; 6: 1–11. https://doi.org/10.1038/s41598-016-0001-8
6. Burby PE, Simmons LA. MutS2 promotes homologous recombination in *Bacillus subtilis*. J Bacteriol. 2017; 199. https://doi.org/10.1128/JB.00682-16 PMID: 27799325
7. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, et al. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell. Elsevier Inc.; 2016; 165: 1493–1506. https://doi.org/10.1016/j.cell.2016.05.003 PMID: 27238023
8. Bryan PN. Protein engineering of subtilisin. Biochim Biophys Acta. 2000; 1543: 203–222. PMID: 11150607
9. Kobayashi T, Kageyama Y, Sumitomo N, Saeki K, Shirai T, Ito S. Contribution of a salt bridge triad to the thermostability of a highly alkaline protease from an alkaliphilic *Bacillus* strain. World J Microbiol Biotechnol. 2005; 21: 961–967. https://doi.org/10.1007/s11274-004-7162-5
10. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence- Specific Control of Gene Expression NIH Public Access. Cell. 2013; 152: 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022 PMID: 23452860
11. Britton RA, Eichenberger P, Gonzalez-Pastor JE, Fawcett P, Monson R, Losick R, et al. Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*. J Bacteriol. American Society for Microbiology (ASM); 2002; 184: 4881–90. https://doi.org/10.1128/JB.184.17.4881-4890.2002 PMID: 12169614
12. Sambrook J, Russell RW. Molecular Cloning: A laboratory manual. 2001.
13. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. Nature Publishing Group; 2009; 6: 343–345. https://doi.org/10.1038/nmeth.1318 PMID: 19363495
14. Che A, Knight T, Canton B, Kelly J, Shetty R. United States Patent No. US 8,999,679 B2. United States of America; 2015. pp. 1–54.
15. Liu H, Wei Z, Dominguez A, Li Y, Wang X, Qi LS. CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation: Fig 1. Bioinformatics. 2015; 31: 3676–3678. https://doi.org/10.1093/bioinformatics/btv423 PMID: 26209430
16. Bryksin A V, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques. NIH Public Access; 2010; 48: 463–5. https://doi.org/10.2144/000113418 PMID: 20569222
17. Anagnostopoulos C, Spizizen J. Requirements for transformation in *Bacillus subtilis*. J Bacteriol. 1961; 81: 741–746. PMID: 16561900
18. Zhang X-Z, Yan X, Cui Z-L, Hong Q, Li S-P. *mazF*, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. Nucleic Acids Res. Oxford University Press; 2006; 34: e71. https://doi.org/10.1093/nar/gkl358 PMID: 16714443

19. Giuliani SE, Frank AM, Collart FR. Functional Assignment of Solute-Binding Proteins of ABC Transporters Using a Fluorescence-Based Thermal Shift Assay. Biochemistry. 2008; 47: 13974–13984. https://doi.org/10.1021/bi801648r PMID: 19063603

20. Cupp-Enyard C. Sigma's Non-specific Protease Activity Assay—Casein as a Substrate. J Vis Exp. MyJoVE Corporation; 2008; https://doi.org/10.3791/899 PMID: 19066538

21. Jain SC, Shinde U, Li Y, Inouye M, Berman HM. The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 å resolution. J Mol Biol. 1998; 284: 137–144. https://doi.org/10.1006/jmbi.1998.2161 PMID: 9811547

22. Shirai T, Suzuki a, Yamane T, Ashida T, Kobayashi T, Hitomi J, et al. High-resolution crystal structure of M-protease: phylogeny aided analysis of the high-alkaline adaptation mechanism. Protein Eng. 1997; 10: 627–634. https://doi.org/10.1093/protein/10.6.627 PMID: 9278275

23. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis. 1997; 18: 2714–2723. https://doi.org/10.1002/elps.1150181505 PMID: 9504803

24. Erwin CR, Barnett BL, Oliver JD, Sullivan JF. Effects of engineered salt bridges on the stability of subtilisin BPN′. Protein Eng Des Sel. 1990; 4: 87–97. https://doi.org/10.1093/protein/4.1.87

Appendix F

RT qRT-PCR results.

Table F.1 – Mean and standard deviation (SD) of the ΔC_q expression of *gfpmut3* in BAC0295 (dCas9-ω(Bs) targeted 16 bp upstream of the -35 promoter element for gfpmut3) and BAC0297 (dCas9-ω(Bs) non-targeting control). The C_q values were averaged from two technical replicates for each of three biological replicates. The ΔC_q values correspond to the difference between the C_q of the gene of interest and the control gene – either *gapA* or *sdhA*.

		ΔC_q Expression	
	Control gene	Mean	SD
BAC0295, *gfpmut3*	*gapA*	1.612	0.275
	sdhA	-1.012	0.554
BAC0297, *gfpmut3*	*gapA*	-3.163	0.238
	sdhA	-6.553	0.675

Table F.2 – Mean and standard deviation (SD) of the ΔC_q expression of *gfpmut3* in BAC0303 (dCas9-ω(Bs) targeted 163 bp upstream of the -35 promoter element for gfpmut3) and BAC0297 (dCas9-ω(Bs) non-targeting control). The C_q values were averaged from two technical replicates for each of three biological replicates. The ΔC_q values correspond to the difference between the C_q of the gene of interest and the control gene – either *gapA* or *sdhA*.

		ΔC_q Expression	
	Control gene	Mean	SD
BAC0303, *gfpmut3*	*gapA*	-3.982	0.245
	sdhA	-5.357	0.191
BAC0297, *gfpmut3*	*gapA*	-3.737	0.344
	sdhA	-5.152	0.494

Appendix G

Copy of “Expanding and understanding the CRISPR toolbox for *Bacillus subtilis* with MAD7 and dMAD7” by Marcus A. Price, Rita Cruz, James Bryson, Franck Escalettes and Susan J. Rosser. Published on 19th February 2020 in Biotechnology and Bioengineering.

Received: 23 October 2019 | Revised: 15 February 2020 | Accepted: 19 February 2020

DOI: 10.1002/bit.27312

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Expanding and understanding the CRISPR toolbox for *Bacillus subtilis* with MAD7 and dMAD7

Marcus A. Price[1] | Rita Cruz[2] | James Bryson[3] | Franck Escalettes[2] | Susan J. Rosser[1,3,4]

[1]School of Biological Sciences, Institute of Quantitative Biology, Biochemistry, and Biotechnology, University of Edinburgh, Edinburgh, UK

[2]Molecular Biology Department, Ingenza Ltd., Roslin Innovation Centre, Roslin, UK

[3]School of Biological Sciences, UK Centre for Mammalian Synthetic Biology, University of Edinburgh, Edinburgh, UK

[4]Centre for Synthetic and Systems Biology and UK Centre for Mammalian Synthetic Biology, School of Biological Sciences, University of Edinburgh, UK

Correspondence
Marcus A. Price, and Susan J. Rosser, School of Biological Sciences, Institute of Quantitative Biology, Biochemistry, and Biotechnology, University of Edinburgh, Edinburgh, UK.
Email: M.A.Price-4@sms.ed.ac.uk (M. A. P.) and Susan.Rosser@ed.ac.uk (S. J. R.)

Funding information
The Industrial Biotechnology Innovation Center (IBioIC), Grant/Award Number: PhD Studentship for M.A.P.

Abstract

The CRISPR-Cas9 system has become increasingly popular for genome engineering across all fields of biological research, including in the Gram-positive model organism *Bacillus subtilis*. A major drawback for the commercial use of Cas9 is the IP landscape requiring a license for its use, as well as reach-through royalties on the final product. Recently an alternative CRISPR nuclease, free to use for industrial R&D, MAD7 was released by Inscripta (CO). Here we report the first use of MAD7 for gene editing in *B. subtilis*, in which editing rates of 93% and 100% were established. Additionally, we engineer the first reported catalytically inactive MAD7 (dMAD7) variant (D877A, E962A, and D1213A) and demonstrate its utility for CRISPR interference (CRISPRi) at up to 71.3% reduction of expression at single and multiplexed target sites within *B. subtilis*. We also confirm the CRISPR-based editing mode of action in *B. subtilis* providing evidence that the nuclease-mediated DNA double-strand break acts as a counterselection mechanism after homologous recombination of the donor DNA.

KEYWORDS

Bacillus subtilis, Cas9, CRISPR, homologous recombination, MAD7

1 | INTRODUCTION

In recent years, precise genome editing with clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) systems have become widely used in many fields of biology (Jinek et al., 2012), enabling significant advances in genome editing tools for industrially relevant microorganisms, such as *Bacillus subtilis* (Altenbuchner, 2016; Burby & Simmons, 2017; Price, Cruz, Baxter, Escalettes, & Rosser, 2019; Westbrook, Moo-Young, & Chou, 2016b). Since the adaptation of the type II CRISPR-Cas9 system from *Streptococcus pyogenes* in 2012 for genomic engineering, along with subsequent iterations—including but not limited to CRISPR interference (CRISPRi) and CRISPR activation systems—it has become the most broadly utilized CRISPR based system in prokaryotes and eukaryotes (Altenbuchner, 2016; Burby & Simmons, 2017; Dicarlo et al., 2013; Jakočiūnas, Jensen, & Keasling, 2015; Jakutyte-Giraitiene & Gasiunas, 2016; Jinek et al., 2012; Lu et al., 2019; Peng et al., 2017; Peters et al., 2016; Price et al., 2019; Westbrook et al., 2016b; K. Zhang, Duan, & Wu, 2016; Zhu et al., 2017). Cas9 induces a blunt DNA double-strand break (DSB) when in complex with either a two-component crRNA–tracrRNA or where these are combined into a single guide RNA complex (Jinek et al., 2012). Previous literature describes that once the DSB is introduced, it can be repaired by nonhomologous end-joining (NHEJ), or by homology directed repair (HDR) when a donor template DNA (dDNA) is supplied (Adli, 2018; Altenbuchner, 2016; Burby & Simmons, 2017; Jinek et al., 2012; Westbrook, Moo-Young, & Chou, 2016a). The inefficiency or total lack of an NHEJ system within most bacteria limits the choice for repair of the cut to HDR in most of these hosts (Shuman & Glickman, 2007).

CRISPRi in bacteria functions through the targeting of a catalytically inactive (D10A and H840A) Cas9 variant, dCas9 to the promoter or within the 5′ region of a gene of interest. This sterically hinders transcription by the RNA polymerase, thus lowering the successful expression of the target gene (Qi et al., 2013). CRISPRi has been exemplified within the Gram-positive model organism *Bacillus subtilis* (Westbrook et al., 2016b), perhaps most notably by Peters et al. (2006) for the functional analysis of all essential genes.

An alternative CRISPR nuclease family, Cpf1 (also known as Cas12a), has similarly been used for genome editing since the first report in 2015 (Zetsche et al., 2015). Cpf1 nucleases exhibit different characteristics to Cas9 nucleases, such as a staggered DSB, a T-rich PAM and the native use of only 1 guide RNA molecule to form a complex with Cpf1 and target the DNA. These characteristics enable Cpf1 nucleases to be used in target organisms or regions within an organism's genome were a lower GC content makes the use of Cas9 less feasible.

While the commercial application of Cas9 nucleases, and increasingly also Cpf1 nucleases, have been widely pursued, a significant drawback for the use of these nucleases is the requirement of a research license and potentially subsequent royalty fees for the commercial exploitation of any developed product.

Recently, Inscripta (CO) released the alternative CRISPR nuclease MAD7 which is free for all commercial or academic research with no reach-through royalties or costs provided the final engineered strain does not contain the MAD7 nuclease (Inscripta, 2019b). As such, its use for commercial genome editing is of great interest. Inscripta report that MAD7 was developed from *Eubacterium rectale* and has proven its functionality in *E. coli*, *S. cerevisiae* and in the human HEK293T cell line. Recently, MAD7 (also known as ErCas12a) was shown to be compatible with genome editing in Zebrafish (Wierson et al., 2019). MAD7 has 31% identity with *Acidaminococcus sp.* BV3L6 Cpf1 (AsCpf1), to which it also shares a T-rich PAM site (5′-YTTN-3′), and a protospacer (the region of the gRNA which associates the nuclease to the DNA target) length of 21 nucleotides (Inscripta, 2019a). A catalytically inactive variant of MAD7 has the potential to be combined with inactive dCas9 and/or ddCpf1 based tools to enable the construction of increasingly sophisticated synthetic biology genetic circuits.

Several CRISPR genome modification systems have been reported for use in the Gram-positive model organism *Bacillus subtilis* based around the efficient homologous recombination (HR) machinery, all of which utilize the Cas9 nuclease (Altenbuchner, 2016; Burby & Simmons, 2017; Price et al., 2019; Westbrook et al., 2016b; K. Zhang et al., 2016), or recently also the partially inactivate nCas9 (D10A) (Liu et al., 2019). Here we demonstrate that CRISPR-Cas9 genome editing in *B. subtilis* 168 is driven primarily by HR by curing *B. subtilis* 168 tryptophan auxotrophy. Subsequently, we show editing with MAD7 is also driven primarily by HR. We used the reporter proteins, AmyE and GFPmut3, to analyze the editing efficiency of this CRISPR-MAD7 system and engineered MAD7 to generate the first reported catalytically inactive, "dead," MAD7 (dMAD7) for single target and multiplexed transcriptional downregulation by dMAD7-mediated CRISPRi.

2 | MATERIALS AND METHODS

2.1 | Strains and media

The strains and plasmids used in this study are listed in Table S1. All oligonucleotides used in this study are listed in Table S2. *Escherichia coli* Top10 cells were used to construct recombinant plasmids. Bacterial cells were cultured in Lysogeny broth (LB) broth or LB agar (VWR) media at 37°C with agitation (250 rpm) where appropriate. *B. subtilis* 168 tryptophan auxotrophy or prototrophy were selected for by growth on M9 agar supplemented with or without 20 μg/mL tryptophan (Harwood & Cutting, 1990). When required, the following antibiotics were supplied to the media: ampicillin (200 μg/mL), chloramphenicol (10 μg/mL), and kanamycin (*E. coli*: 100 μg/mL; *B. subtilis*: 10 μg/mL).

2.2 | Plasmid and strain construction

Unless otherwise stated, plasmid construction was performed as described in Sambrook and Russell (2001). DNA oligonucleotides were purchased from Merck or Integrated DNA Technologies. The reagents for polymerase chain reaction (PCR), restriction digest, DNA phosphorylation, and ligation were purchased from New England Biolabs (NEB). DNA purification was performed using the Monarch® Nucleic Acid Purification Kits from NEB. DNA sequences were confirmed by Source Bioscience.

2.2.1 | CRISPR plasmid construction

pBAC0155 was constructed using the inABLE plasmid assembly method (Che, Knight, Canton, Kelly, & Shetty, 2015; Price et al., 2019). Individual 5′ truncated parts were prepared by PCR followed by restriction digest at 5′ and 3′ regions with SapI. Where parts <120 base pair (bp) were used, complementary oligonucleotides were phosphorylated and annealed leaving three nucleotides (nts) single strands at both the 5′ and 3′ ends to remove the need for SapI digestion. The parts were ligated to phosphorylated and annealed oligonucleotides at each terminus, containing 3 nt and 16 nt single strands at the 5′ and 3′ ends, respectively. The part- oligonucleotides fusions were annealed at the homologous 16 nt overhangs for 1 hr at 37°C, and used to transform electrocompetent *E. coli*.

pBAC0155 consisted of four parts: 1. the *E. coli/B. subtilis* shuttle vector backbone from pBAC0001; 2. the LacI repressor and isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible P_{grac} promoter from pBAC0001, including a multiple cloning site; 3. the bidirectional strong *rrnB* T1 and T2 terminators; 4. the MAD7 gRNA expression module consisting of the P_{veg} promoter, the "MAD7 handle" section of the gRNA and spacer DNA flanked by SapI sites for cloning of the protospacer DNA. Three 5′ truncated parts were prepared by PCR from the indicated template and oligonucleotides: 1. pBAC0001 with oMAP0010/0011; 2. pBAC0001 with oMAP0018/0019; 3. pING0001

with oMAP0024/0025. The short part 4 was prepared as described from oMAP0492/0493. Parts were ligated at 5′ and 3′ respectively with annealed oligonucleotides: 1. oMAP0008/0009 and oMAP0014/0015; 2. oMAP0016/0017 and oMAP0020/0021; 3. oMAP0022/0023 and oMAP0486/0487; 4. oMAP0488/0489 and oMAP0498/0499.

pBAC0158 was constructed by digesting pBAC0155 and pMK-RQ-MAD7 (*MAD7*, codon-optimized for *B. subtilis* and flanked by BamHI and XbaI recognition sites, synthesized by Thermo Fisher Scientific) with BamHI and XbaI restriction enzymes, and ligating the pBAC0155 backbone with the *MAD7* gene using T4 DNA ligase.

pBAC0184, pBAC0194, and pBAC0195 were constructed by PCR amplifying dCas9 from pdCas9-bacteria with oligonucleotides oMAP0073/0074, introducing a BsaI site and XbaI recognition sites at the 5′ and 3′ end of the gene, respectively. The amplified *dcas9* was digested with BsaI and XbaI. pBAC0041, pBAC0035, and pBAC0165 were digested with BamHI and XbaI to remove the *cas9* gene. The *dcas9* gene fragment was subsequently ligated with the pBAC0041/0035/0165 backbones using T4 DNA ligase. pdCas9-bacteria was a gift from Stanley Qi (plasmid #44249; Addgene; Qi et al., 2013).

pBAC0187, pBAC0188, pBAC0189, and pBAC0190 were constructed by digesting pBAC0158, pBAC0162, pBAC0163, and pBAC0166, respectively, as well as pMK-RQ-dMAD7 (*dMAD7* [*MAD7* gene with D877A [codon GCT to GAT], E962A [codon GCA to GAA], D1213A [codon GCT to GAT] modifications] flanked by BamHI and XbaI recognition sites, synthesised by Thermo Fisher Scientific), with BamHI and XbaI. The *dMAD7* gene fragment was subsequently ligated with the pBAC0158/0162/0163/0166 backbones using T4 DNA ligase.

The protospacer insertion into pBAC0015, for the completion of Cas9-mediated genome editing plasmids, was carried out as previously described (Price et al., 2019). The construction of CRISPR plasmids for MAD7 or dMAD7 was carried out in a similar fashion, however SapI, instead of AarI, was used to cleave the pBAC0158 and pBAC0187 backbones to yield 3 nt single-stranded DNA overhangs compatible with the gRNA protospacer constructed by annealing of an oligonucleotide pair. All protospacer regions were identified using the online tool, CHOPCHOP (Labun, Montague, Gagnon, Thyme, & Valen, 2016; Montague, Cruz, Gagnon, Church, & Valen, 2014).

The following plasmids for Cas9-mediated editing were prepared with the pBAC0015 backbone using the indicated oligonucleotide pairs: pBAC0129 (oMAP0386/0387); pBAC0165 (oMAP0573/0574); pBAC0185 (oMAP0694/0695). The following plasmids for MAD7-mediated editing were prepared from the pBAC00158 backbone using the indicated oligonucleotide pairs: pBAC0162 (oMAP0549/0550); pBAC0163 (oMAP0553/0554); pBAC0166 (oMAP0571/0572); pBAC0218 (oMAP0799/0800). The following plasmids where prepared from the pBAC0187 backbone using the indicated oligonucleotide pairs: pBAC0207 (oMAP0759/0760); pBAC0208 (oMAP0761/0762); pBAC0209 (oMAP0763/0764); pBAC0210 (oMAP0765/0766); pBAC0211 (oMAP0767/0768); pBAC0212 (oMAP0769/0770); pBAC0213 (oMAP0771/0772); pBAC0214 (oMAP0775/0776); pBAC0215 (oMAP0777/0778); pBAC0219 (oMAP0793/0794); pBAC0220 (oMAP0797/0798); and pBAC0222 (oMAP0795/0796).

Schematic representations of plasmids pBAC0015, pBAC0155, pBAC0158, and pBAC0187 can be found in the supplementary information.

2.2.2 | dDNA preparation

dDNA, encoding a programmed target site edit together with a synonymous PAM mutation to eliminate Cas9 or MAD7 cleavage at the edited site after HR or HDR was constructed by overlap extension PCR (OE-PCR) as described previously (Bryksin & Matsumura, 2010; Price et al., 2019).

2.2.3 | Strain construction

BAC0111 was constructed by transforming naturally competent *B. subtilis* 168 with the integration plasmid pGFPbgIS (Anagnostopoulos & Spizizen, 1961; Bennallack, Burt, Heder, Robison, & Griffitts, 2014; Bisicchia, Botella, & Devine, 2010). Transformants were selected on LB agar plates supplemented with kanamycin (Figure S1).

A total of 1 µg of an OE-PCR product generated using the oligonucleotides set oMAP0388/0393/0394/0395 and the genomic DNA (gDNA) of strain BAC0111 as template was co-transformed alongside 200 ng of the editing plasmid pBAC0129 for the modification of BAC0111 to insert P_{veg} upstream of *gfpmut3*. The editing plasmid was removed from the edited strain by overnight growth in LB supplemented with 1 mmol/L IPTG and subsequent confirmation by counter plating on LB agar plates with and without chloramphenicol, yielding strain BAC0288 (Figure S1; Price et al., 2019). The insertion of P_{veg} was verified by sequencing and by fluorescence emission analysis using Safe Imager 2.0 Blue Light Transilluminator and Amber Filter System (excitation: 470 nm, emission: 530 nm; Thermo Fisher Scientific) to detect GFPmut3.

BAC0348-0355 and BAC0360-0368 were constructed by transforming naturally competent cells of the relevant parental strain with the appropriate plasmid (Table S1; Anagnostopoulos & Spizizen, 1961; Bennallack et al., 2014). Transformants were selected on LB agar plates supplemented with chloramphenicol.

2.3 | CRISPR-Cas9/MAD7-Mediated gene editing in *B. subtilis*

A single OE-PCR product was used to allow a direct comparison between Cas9 and MAD7 editing efficiencies at the *amyE* locus. The OE-PCR product generated using the oligonucleotide set oMAP0121/0551/0552/0122 was co-transformed alongside the editing plasmids pBAC0041 and pBAC0162.

dDNA for *gfpmut3* editing was generated by PCR using oligos oMAP0575/0578 and the synthesized plasmid pMK-RQ-gfpmut3-dDNA as a template.

Genome editing was carried out in triplicate by co-transforming naturally competent *B. subtilis* with 200 ng editing plasmid DNA and 1 µg

dDNA (Anagnostopoulos & Spizizen, 1961; Bennallack et al., 2014; Price et al., 2019). When targeting *gfpmut3*, transformants were spread on LB agar plates supplemented with chloramphenicol and IPTG (1 mmol/L). Effective knock-out of *gfpmut3* by stop codon introduction was determined by analysis of fluorescence emission using Safe Imager 2.0 Blue Light Transilluminator and Amber Filter System (excitation: 470 nm, emission: 530 nm; Thermo Fisher Scientific). Genotypes were confirmed by colony PCR with oligonucleotides hybridizing to the chromosome (oMAP0393/0814) outside of the dDNA homology arm region and sequenced with primer oMAP0815 to ensure accurate coverage of the targeted region. When targeting *amyE*, transformants were spread on LB agar plates supplemented with chloramphenicol, IPTG (1 mmol/L) and 1% soluble potato starch (VWR). Effective knock-out of *amyE* by stop codon introduction was determined by staining transformation plates with iodine (Price et al., 2019; Zhang, Yan, Cui, Hong, & Li, 2006). Genotypes were confirmed by colony PCR with oligonucleotides hybridizing to the chromosome (oMAP0811/0812) outside of the dDNA homology arm region and sequenced with primer oMAP0813 to ensure accurate coverage of the targeted region.

2.4 | Restoring *B. subtilis* tryptophan prototrophy

The CRISPR-Cas9 genome editing mechanism in *B. subtilis* was investigated by restoring tryptophan prototrophy when transforming an OE-PCR product (oligonucleotides set oMAP0236/0696/0697/0239 and gDNA of *B. subtilis* 168 as template) containing the mutation needed to remove the *trpC2* lesion, in the presence and absence of editing plasmid pBAC0185, or Nontargeting plasmid pBAC0035 (Altenbuchner, 2016). Transformations were carried out in triplicate in naturally competent *B. subtilis* as previously described, and tryptophan prototroph cells were selected in M9 minimal medium (Anagnostopoulos & Spizizen, 1961; Bennallack et al., 2014; Price et al., 2019). Before spreading transformants on M9 agar plates, the cells were washed three times with 10 mmol/L phosphate-buffered saline to ensure there was no carryover of tryptophan from the transformation process.

The CRISPR-MAD7 genome editing mechanism in *B. subtilis* was investigated in the same manner. OE-PCR product (oligonucleotides set oMAP0236/0801/0802/0239 and gDNA of *B. subtilis* 168 as template) containing the mutation needed to remove the *trpC2* lesion was transformed in the presence and absence of editing plasmid pBAC0218, or nontargeting plasmid pBAC0163. IPTG was included in all transformation plates to limit the background of nonselected colony forming units (CFUs).

Transformation results between different batches of competent cells were normalized by the transformation efficiency obtained when transforming only the nontargeting plasmid for the nuclease being analyzed and selecting on LB agar supplemented with chloramphenicol (Cas9), or LB agar supplemented with chloramphenicol and IPTG (MAD7).

Unpaired *t* tests with Welch's correction were performed to determine two-tailed *p* values and identify statistically significant or nonsignificant differences between the number of colonies obtained.

2.5 | Quantitative analysis of α-amylase activity

Relative extracellular α-amylase activity was quantified in the culture supernatant of strains BAC0352-0355 and BAC0360-0363 using a starch degradation assay. For each strain, an overnight culture was used to inoculate a pre-culture supplemented with chloramphenicol and 1 mmol/L IPTG. Once in exponential growth phase, the pre-culture was back diluted in triplicate into fresh and prewarmed medium also supplemented with chloramphenicol and 1 mmol/L IPTG and grown for 24 hr. Supernatant samples were clarified by centrifugation and 25 µL were mixed in triplicate with 100 µL assay solution (50 mmol/L Tris-HCl, pH 6.8, 25 mmol/L $CaCl_2$, 0.05% [wt/vol] soluble potato starch) and incubated for 30 min at 37°C. 50 µL stop solution (1 mol/L HCl, 0.01% [wt/vol] I_2, 0.1% [wt/vol] KI) was added and absorbance at 620 nm measured. Unpaired *t* tests with Welch's correction were performed to determine two-tailed *p* values and identify statistically significant or nonsignificant differences.

2.6 | Quantitative analysis of GFPmut3 expression detection

Relative fluorescence was quantified in the cultures of strains BAC0348-0351 and BAC0364-0368. For each strain, an overnight culture was used to inoculate a pre-culture supplemented with chloramphenicol and 1 mmol/L IPTG. Once in exponential growth phase, the pre-culture was back diluted into fresh and prewarmed medium also supplemented with chloramphenicol and 1 mmol/L IPTG and grown for 24 hr. The culture fluorescence was measured in samples of 100 µL using the FLUOstar Omega Microplate Reader (BMG LABTECH) in 96-well flat-bottom plates (excitation: 485 nm; emission: 520 nm; gain: 1,000; Greiner). Unpaired *t* tests with Welch's correction were performed to determine two-tailed *p* values and identify statistically significant or nonsignificant differences.

3 | RESULTS

3.1 | Comparison of CRISPR-Cas9 and CRISPR-MAD7 enabled genome editing efficiencies

To compare the gene-editing efficiencies between Cas9 and MAD7 in *B. subtilis*, chromosomally expressed *amyE* and *gfpmut3* genes were selected due to ease of analysis of successfully edited colonies by starch degradation or fluorescence respectively. For both nucleases, a single dDNA with 1 kbp homology arms, either side of the site targeted for modification, was designed to introduce a stop codon at the 5′ of the gene and remove the PAM sites selected based on their proximity (Figure 1a,b).

Cas9 and MAD7 yielded *amyE* knock-out efficiencies of 98% and 93% respectively when the editing plasmids were co-transformed with dDNA to *B. subtilis* 168 (Figure 1c). When knocking out *gfpmut3*, editing efficiencies of 75% and 100% were observed for Cas9 and

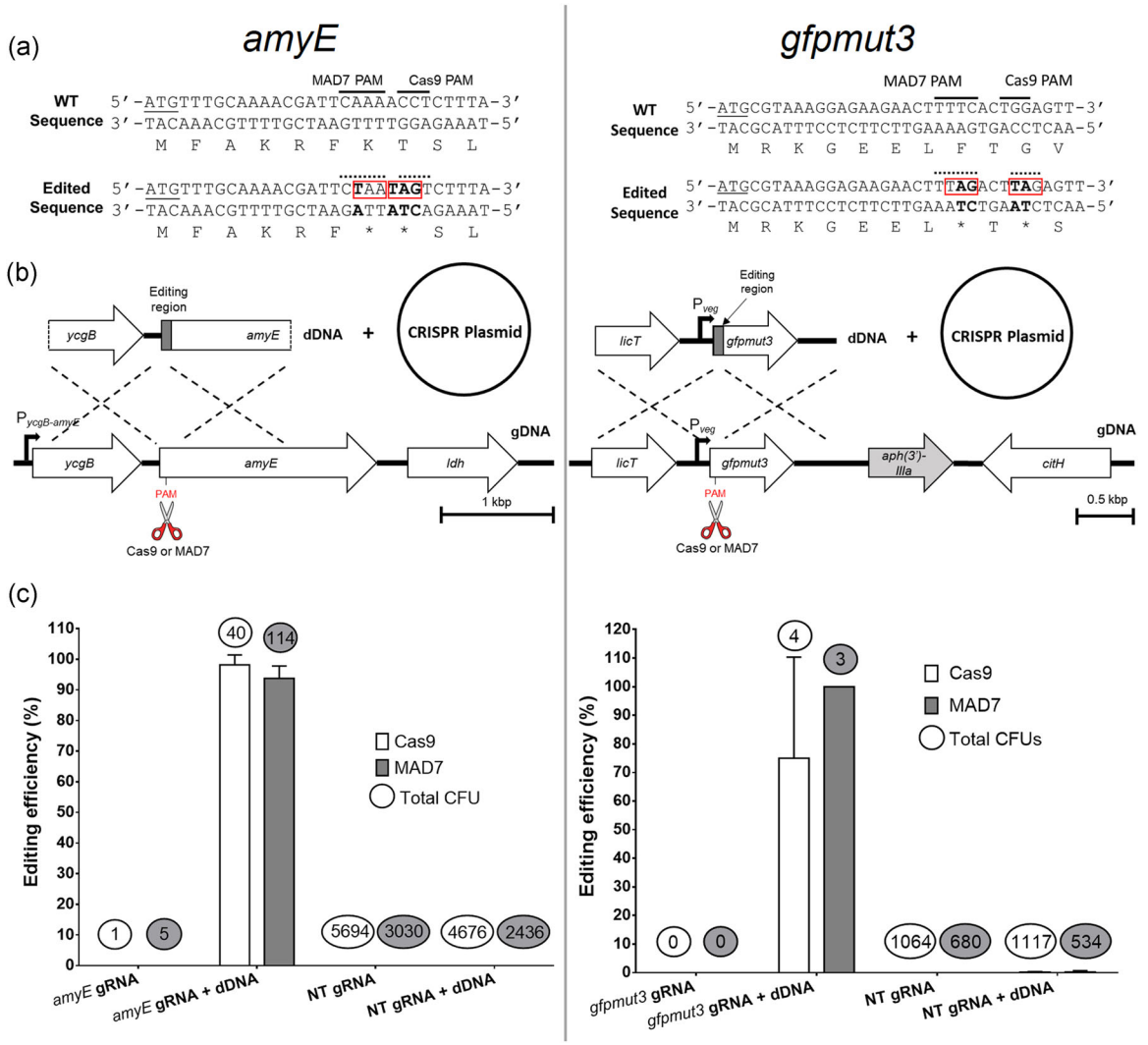


FIGURE 1 CRISPR-Cas9 or CRISPR-MAD7-mediated editing of *amyE* and *gfpmut3*. (a) Non-edited (WT) and edited sequences with their corresponding amino acid sequences (*=stop codon). The targeted PAM sites are indicated for both Cas9 and MAD7. The modified base pairs are highlighted in bold and the introduced stop codons are marked with red boxes. (b) Co-transformational editing approach were the CRISPR plasmid expressing the gRNA and nuclease is transformed alongside a linear editing template (dDNA) containing the editing region. (c) Editing efficiency following co-transformation of *B. subtilis* 168 (*amyE*) or BAC0288 (*gfpmut3*). Bars represent the average editing efficiency obtained following transformation of the targeting or nontargeting (NT) gRNA expression plasmids for each nuclease with or without dDNA (*amyE* targeting: Cas9: pBAC0041; MAD7: pBAC0162. *gfpmut3*: Cas9: pBAC0165; MAD7: pBAC0166. NT: Cas9: pBAC0035; MAD7: pBAC0163). Editing efficiency (%) was determined by observing either starch degradation or fluorescence in the transformation plates. The circled number above each bar indicates the total number of colonies phenotypically screened. Error bars indicate the standard deviation between three independent transformation events. dRNA, donor template DNA; gRNA, guide RNA; WT, wild-type [Color figure can be viewed at wileyonlinelibrary.com]

MAD7 respectively despite the lower transformation efficiency of BAC0288 (Figure 1c). In the absence of dDNA, the DSB catalyzed by either nuclease drastically reduces cell viability and no successfully edited colonies were identified. Following phenotypic analysis, the genotypes of a selected population of the transformants were confirmed by colony PCR and sequencing for both *amyE* and *gfpmut3* (Table S3).

Nontargeting plasmids, containing a gRNA with a random DNA sequence which does not target anywhere in the genome of *B. subtilis* 168 were transformed with and without dDNA for both *amyE* and *gfpmut3* to determine the basal level of editing in the absence of the DSB induced by a CRISPR nuclease (Figure 1c). A basal level of editing of 2 in 1,117 CFUs and 1 in 534 CFUs was phenotypically detected when the dDNA for *gfpmut3* editing was co-transformed with the Cas9 or MAD7 nontargeting plasmids, respectively. In the case of *amyE*, no edited colonies were detected with the same strategy. This highlights the importance of having both the nuclease and targeting gRNA present to ensure high editing efficiency is obtained due to the counter-selection of unedited colonies. The higher number of colonies analyzed following transformation of the nontargeting plasmids was due to the absence of the selective pressure, against colony formation, in non-edited cells when a targeting gRNA is present.

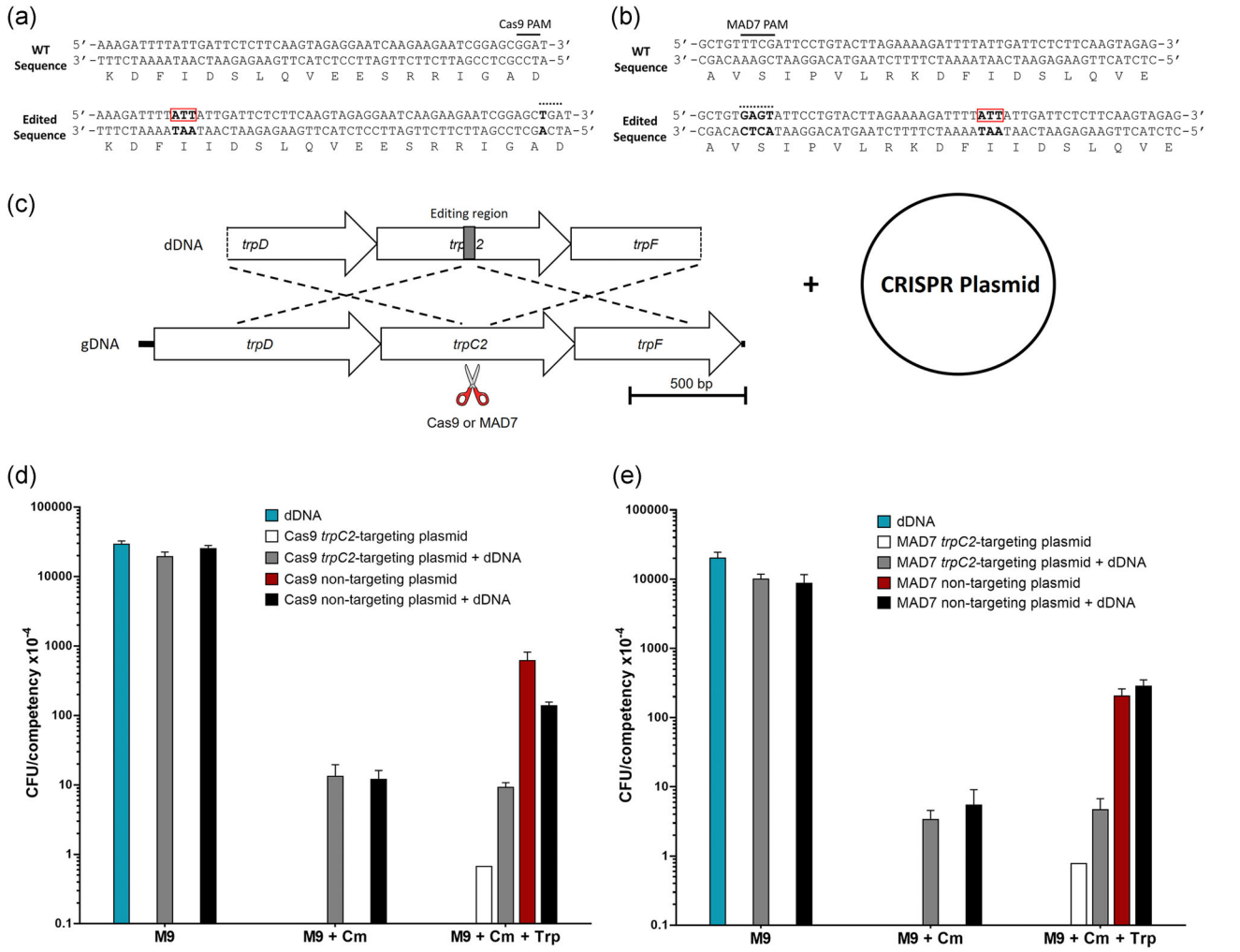


FIGURE 2 Restoration of *B. subtilis* prototrophy using CRISPR-Cas9 and CRISPR-MAD7 for genome editing. (a and b) show the non-edited (WT) and edited sequences for Cas9 and MAD7 editing respectively as well as their corresponding amino acid sequences. The modified base pairs are highlighted in bold and the inserted isoleucine (Ile) codon, adjacent to I110, is marked with red boxes. (c) Co-transformational editing approach were the CRISPR plasmid expressing the gRNA and nuclease is transformed alongside a linear editing template (dDNA) containing the editing region. (d and e) Graphs show the number of transformants following transformations with the indicated combinations of dDNA and targeting (pBAC0185 for Cas9 and pBAC0218 for MAD7) or nontargeting plasmid (pBAC0035 for Cas9 and pBAC0163 for MAD7) to restore *B. subtilis* 168 prototrophy, with Cas9 and MAD7, respectively. Bars represent the average number of colony-forming unit (CFU) normalized by the transformation efficiency of pBAC0035 (Cas9) selected on LB agar supplemented with chloramphenicol or pBAC0163 (MAD7) selected on LB agar supplemented with chloramphenicol and IPTG. Error bars indicate the standard deviation between three independent transformation events. Cm, chloramphenicol; dRNA, donor template DNA; IPTG, isopropyl β-D-1-thiogalactopyranoside; LB, Lysogeny broth; Trp, tryptophan; WT, wild-type [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | Homologous recombination versus DNA double-strand break repair as the driving mechanism for CRISPR genome editing in *B. subtilis* 168

To elucidate the mechanism with which CRISPR-mediated editing takes place in *B. subtilis* 168, we made use of this strain's tryptophan auxotrophy to compare the efficiency in restoring prototrophy when a linear dDNA is transformed on its own, or in the presence of either a Cas9 or MAD7 nontargeting or *trpC2*-targeting plasmid. The linear dDNA, harboring 1 kbp homology regions either side of the site targeted for modification, was designed to simultaneously introduce an additional isoleucine residue adjacent to I110 residue of TrpC2, returning the strain to a prototrophic state, and a synonymous mutation to eliminate the PAM recognition site and prevent continuous cutting by the *trpC2*-targeting plasmid (Figure 2a,b; Altenbuchner, 2016). By selecting transformants in M9 minimal medium supplemented with or without either chloramphenicol (plasmid selection) or tryptophan, we could clarify whether HR drives genome editing, preventing a DSB, or if the DSB induces DNA repair by HR. As the procedure to induce natural competence utilizes tryptophan within the growth medium throughout, there is no selection for prototrophic cells before the spreading of the transformants on the agar plates.

In the absence of tryptophan, there is not a significant difference in CFU obtained when transforming the linear dDNA to restore tryptophan prototrophy on its own or in the presence of either the *trpC2* targeting or nontargeting plasmids (Figure 2). Furthermore, when cells with restored

prototrophy were also selected in the presence of chloramphenicol, there was no significant difference between the co-transformation of dDNA with either the *trpC2*-targeting or nontargeting CRISPR-Cas9 plasmids (Figure 2). Both these results indicate that HR is the main driving force for CRISPR-Cas9 editing in the presence of dDNA. When M9 was supplemented with tryptophan, the absence of selective pressure for restored prototrophy results in a significantly lower number of CFU when co-transforming the dDNA and the *trpC2*-targeting plasmid compared with the co-transformation of dDNA with the nontargeting plasmid (Figure 2). In this case, the lethal cut induced by the nuclease counterselects the transformants in which HR of dDNA has not occurred. As such, while the high efficiency of HR is the main driving force for genome editing, the nuclease induced DSB is essential to obtain high editing efficiency in *B. subtilis* 168.

TABLE 1 Average number of transformants (CFU) obtained following triplicate transformation reactions of naturally competent *B. subtilis* 168 and BAC0288 with 200 ng of Cas9, dCas9, MAD7, and dMAD7 plasmids

Parental strain	Target	Plasmid	Nuclease	Average CFU obtained
B. subtilis 168	*amyE*	pBAC0041	Cas9	0.3
		pBAC0184	dCas9	898
		pBAC0162	MAD7	1.7
		pBAC0188	dMAD7	1,248
BAC0288	*gfpmut3*	pBAC0165	Cas9	0
		pBAC0195	dCas9	392
		pBAC0166	MAD7	0.7
		pBAC0190	dMAD7	343

Abbreviation: CFU, colony-forming unit.

3.3 | Engineering of MAD7 to construct and characterize the catalytically inactive dMAD7

3.3.1 | Identification of MAD7 catalytic residues

A pairwise alignment of the amino acid sequences of MAD7 and AsCpf1, confirmed the 31% identity (Figure S2). Previously, it was reported that the catalytic residues of AsCpf1 are Asp908, Glu993, and Asp1263 (Yamano et al., 2016; Zetsche et al., 2015). Asp908 lies in a region of high similarity with MAD7, with residues 905-916 corresponding to MAD7 residues 874-885. AsCpf1 Glu993 does not lie in a region of high homology, however, the alignment revealed that this residue was conserved in MAD7. Finally, the residue corresponding to Asp1263 in AsCpf1 was found in a region on high homology with AsCpf1, with residues 1261–1268 corresponding to MAD7 residues 1211–1218. The corresponding catalytic residues in MAD7 (Asp877, Glu962, and Asp1213) identified by sequence homology (Figure S2) were simultaneously modified to alanine in silico and the corresponding gene *dMAD7* was synthesized.

3.3.2 | dMAD7 lacks the ability to induce a lethal DNA double-strand break

To verify whether the DNA cleavage capacity of MAD7 was removed in the putative dMAD7, the synthesized *dMAD7* gene was used to replace *MAD7* within the *amyE* and *gfpmut3* targeting plasmids pBAC0162 and pBAC0166, respectively. As a control, the well-characterized *dcas9* was used to replace *cas9* in plasmids pBAC0041 and pBAC0165 (Peters et al., 2016; Qi et al., 2013; Westbrook et al., 2016b). Naturally competent *B. subtilis* 168 and BAC0288 were respectively transformed with the *amyE* and *gfpmut3* targeting plasmids with both active and inactive nuclease variants. Triplicate transformations were spread on plates supplemented with chloramphenicol and IPTG to ensure nuclease expression. The average number of CFUs obtained for each set of transformations (Table 1) indicates that the engineered dMAD7 does not catalyze DSB of DNA since it does not cause the reduced viability observed for the catalytically active nuclease.

3.3.3 | dMAD7 retains DNA binding capacity to enable CRISPRi

Extracellular α-amylase activity was quantified in strains expressing dMAD7 targeting five PAM sites (5′-TTTN-3′) at the 5′ end of *amyE*, two on the template strand and three on the nontemplate strand. The results were directly compared with strains expressing dCas9 targeting *amyE* +25 bases downstream of the start codon. Strains expressing nontargeting dMAD7 and dCas9 plasmids were used as the negative controls for downregulation (Figure 3a,b).

The results confirm that dCas9-mediated CRISPRi is highly efficient with a 99.4% reduction in α-amylase activity, while dMAD7-mediated CRISPRi appears less efficient, ranging from 59.3% to 51.5% activity reduction depending on the gRNA and PAM site sequence. The gRNA targeting the PAM site TTTG +21 bases from the start codon did not exhibit significantly reduced levels of activity. As this PAM sequence was the same as others where CRISPRi was successful, and the GC % of the protospacer (28.6%) is similar to the one targeting the PAM site +4 (33.3%), we hypothesize the cause of this lowered efficiency is due to secondary structure within the gRNA, as has previously been reported for Cas9 (Thyme, Akhmetova, Montague, Valen, & Schier, 2016; Xu, Lian, Jia, Li, & Huang, 2017).

To further investigate the capacity of dMAD7 for CRISPRi, the *gfpmut3* gene in strain BAC0288 was targeted. Here, six PAM sites (three on each strand) where targeted with the 5′-YTTN-3′ PAM sequence recommended by Inscripta (Figure 3c). dMAD7 was targeted to the 5′ end of *gfpmut3* and fluorescence was compared with strains expressing dCas9 targeting *gfpmut3* +27 bases downstream of the start codon. Strains expressing nontargeting dMAD7 and dCas9 plasmids were used as negative controls for

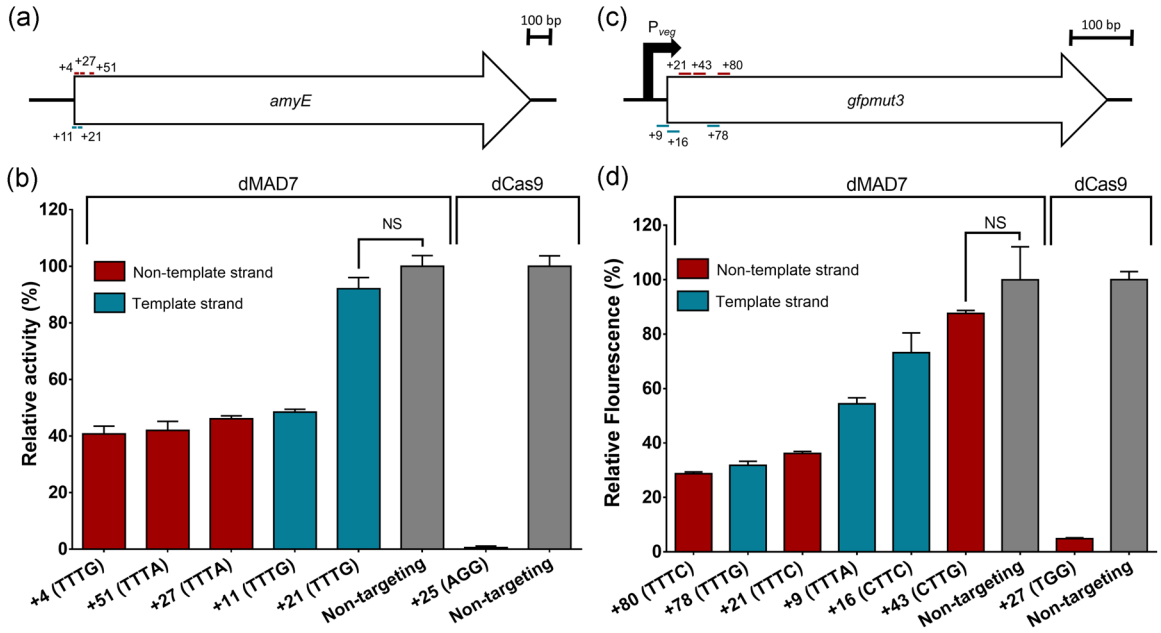


FIGURE 3 Relative α-amylase activity and GFPmut3 fluorescence following dCas9 and dMAD7-mediated CRISPRi. (a and c) Schematic diagrams of gRNA binding sites for dMAD7 within *amyE* and *gfpmut3* respectively. Values represent the distance of each targeted PAM site from the start codon on either the template (blue) or non-template (red) DNA strand. (b) Bar graph represents the extracellular α-amylase activity normalized by $OD_{600\,nm}$ relative to the nontargeting gRNA control for either dMAD7 or dCas9, after 24 hr of growth. The horizontal axis indicates the PAM site targeted by each gRNA for both dMAD7 and dCas9 and its distance to the *amyE* start codon (+4: BAC0360; +51: BAC0363; +27: BAC0362; +11: BAC0361; +21: BAC0355; +25: BAC0353). Red and blue bars correspond to PAM sites on the non-template and template strands, respectively. Gray bars represent nontargeting controls (dMAD7: BAC0354; dCas9: BAC0352). (d) Bar graph represents the fluorescence intensity normalized by $OD_{600\,nm}$ relative to the nontargeting gRNA control for either dMAD7 or dCas9, after 24 hr of growth. The horizontal axis indicates the PAM site targeted by each gRNA for both dMAD7 and dCas9 and its distance to the *gfpmut3* start codon (+80: BAC0368; +78: BAC0367; +21: BAC0351; +9: BAC0364; +16: BAC0365; +43: BAC0366; +27: BAC0349). Red and blue bars correspond to PAM sites on the non-template and template strands, respectively. Gray bars represent the nontargeting controls (dMAD7: BAC0350; dCas9: BAC0348). Error bars indicate the standard deviation between three biological replicates. Two-tailed *p* values following unpaired *t* test with Welch's correction. gRNA, guide RNA; NS, not significant [Color figure can be viewed at wileyonlinelibrary.com]

downregulation (Figure 3d). Here, a broader effect on expression was observed for dMAD7 with statistically significant CRISPRi efficiencies ranging from 71.3% to 26.8%. Once again, the *gfpmut3* targeting dCas9 control exhibited highly efficient CRISPRi with a 95.1% reduction in expression. This difference in efficiency may indicate that further modifications to dMAD7 could be made to increase its ability to bind DNA efficiently.

3.3.4 | Multiplexed CRISPRi of *amyE* and *gfpmut3* with dMAD7

To increase the efficiency of dMAD7 transcriptional downregulation, multiplexed targeting, where more than one gRNA is utilized at a time, was tested. gRNA arrays were inserted in the same manner as single gRNAs, with the final array designed with the dMAD7 handle direct repeat at both the 3′ and 5′ ends, as well as between the inserted protospacers.

Additive downregulation of *amyE* was tested by targeting PAM sites at +4 TTTG and +51 TTTA in strain BAC0377 (Figure 4b). Similarly, *gfpmut3* was analyzed by targeting PAM sites at +80 TTTC and +21 TTTG in strain BAC0380 (Figure 4b). These were compared with the nontargeting dMAD7 control strain BAC0350, as well as the single gRNA, *amyE* or *gfpmut3* targeting dMAD7 strains BAC0381, BAC0382, BAC0351, and BAC0368. To ensure the gRNA array was matured from a single transcript into single gRNA units, strain BAC0378, carrying a gRNA array to target *amyE* +4 TTTG and *gfpmut3* +80 TTTC was analyzed for transcriptional interference of both targets (Figure 4c).

The results show when *amyE* and *gfpmut3* were simultaneously targeted the downregulation of both genes was found to be similar to when only one was targeted for CRISPRi (Figure 4c). As such, the gRNA array was correctly matured to single gRNAs and there is no significant competition between the two gRNAs for dMAD7-mediated CRISPRi.

When two gRNAs are combined to target either *gfpmut3* or *amyE*, the measured downregulation is not cumulative. As the BAC0378 multiplexing results indicate no significant competition between the gRNAs for dMAD7, it is thought that there is a potential steric hindrance between the protospacer-dMAD7 complexes used to simultaneously target each reporter.

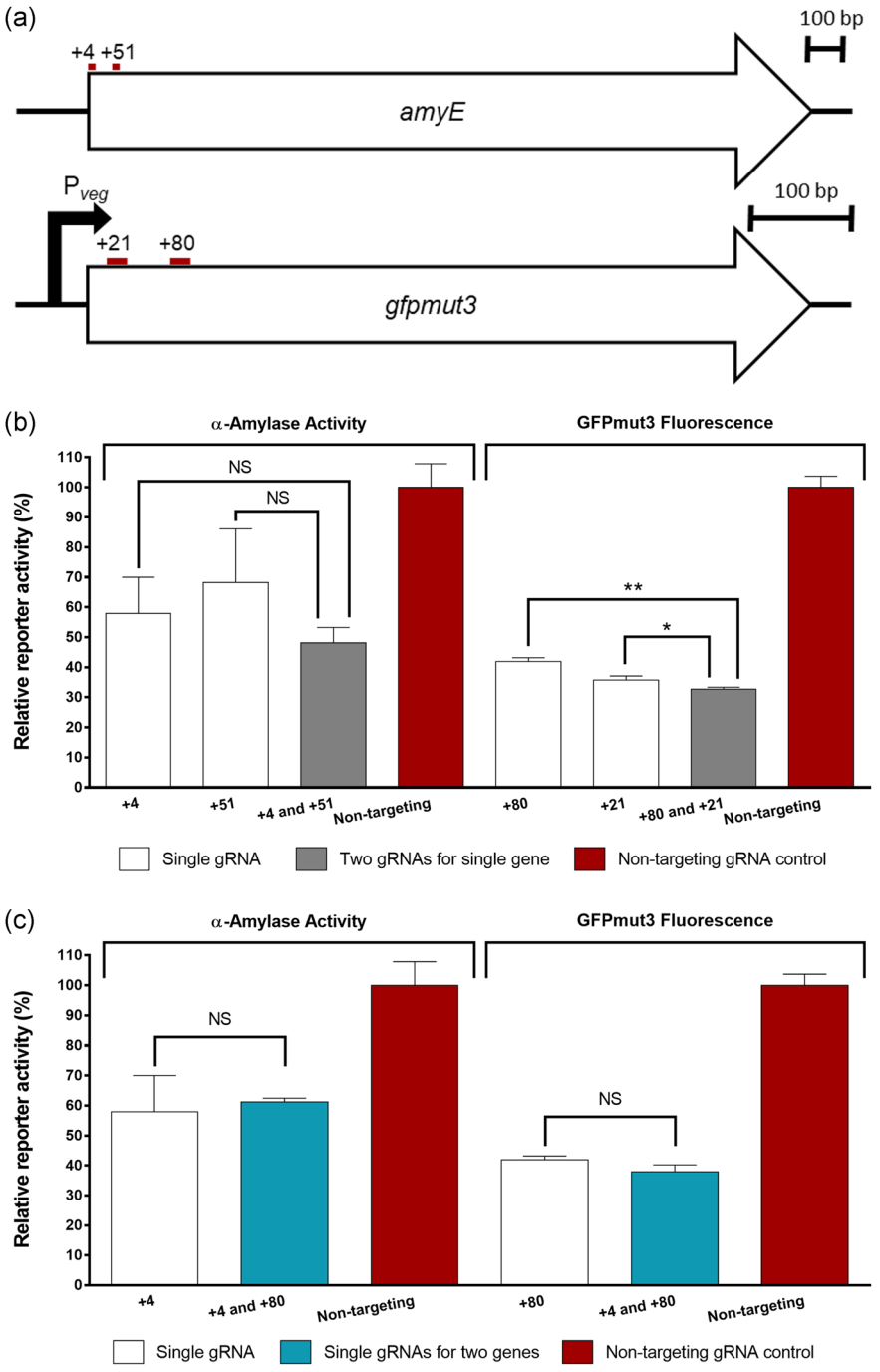


FIGURE 4 Relative α-amylase activity and GFPmut3 fluorescence following dMAD7-mediated multiplexing CRISPRi in BAC0288. (a) Schematic diagrams of gRNA binding sites for dMAD7 within *amyE* and *gfpmut3*. Values represent the distance of each targeted PAM site from the start codon on the non-template DNA strand. (b) Investigation for an additive multiplexed CRISPRi effect when two gRNAs target a single gene. (c) Investigation for a multiplexed CRISPRi effect when two gRNAs target different genes within the same strain. Bars represent extracellular α-amylase activity, or GFPmut3 fluorescence intensity, normalized by $OD_{600\,nm}$, relative to the nontargeting gRNA control strain, after 24 hr of growth. The horizontal axis indicates the targeted PAM site(s) distance to the *amyE* or *gfpmut3* start codon (Strains used: +4: BAC0381; +51: BAC0382;+4 and +51: BAC0377; +4 and +80: BAC0378; +80: BAC0368; +21: BAC0351; +80 and +21: BAC0380). White bars indicate strains in which a single gRNA is utilized to target a single gene. Gray bars indicate strains in which the effect of two gRNAs on a single gene is investigated. Blue bars indicate BAC0378 in which the effect of single gRNAs on two different genes is investigated. Red bars indicate the negative control strain, BAC0350, in which a nontargeting gRNA is expressed alongside *dMAD7*. Error bars indicate the standard deviation between three biological replicates. $^{*}p < .05$; $^{**}p < .01$. Two-tailed *p* values were derived following unpaired *t* test with Welch's correction. gRNA, guide RNA; NS, not significant [Color figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

We have exemplified the first reported use of the CRISPR-associated nuclease MAD7 free for all commercial or academic research in the Gram-positive model organism, *B. subtilis*. The MAD7-mediated genome editing efficiency determined by targeting *amyE* (93%) and *gfpmut3* (100%) was comparable to the commonly used Cas9 nuclease. This indicates that MAD7 is a viable alternative to Cas9 for strain development of the industrial workhorse *B. subtilis*. These results also indicate that the killing efficiency of MAD7 (*amyE*: 99.84%; *trpC2*: 99.62%) in *B. subtilis*, determined by the total number of CFUs following transformation with targeting versus nontargeting plasmids only, was similar to that of Cas9 (*amyE*: 99.98%; *trpC2*: 99.89%). It is hypothesized that, due to the strong selection for survival, these "escaper" colonies may harbor mutations deactivating the killing capacity of the CRISPR systems. Such mutations may occur within the editing plasmid, or on the chromosome at the PAM site or the first 10–12 nt of the gRNA protospacer (known as the seed region) within which any mutations cause a severely deleterious effect on cleavage efficiency (Jiang, Bikard, Cox, Zhang, & Marraffini, 2013; Jinek et al., 2012).

We have provided evidence that, in *B. subtilis* 168, CRISPR-Cas9 and CRISPR-MAD7 genome editing efficiency is driven primarily by HR of dDNA preventing the lethal Cas9 or MAD7-induced DNA DSB, rather than HDR following the DSB. Mougiakos et al. reported the endogenous HR machinery within *Bacillus smithii* incorporating plasmid-borne dDNA while Cas9 was inactive at ≥42°C. Counterselection of the cells which had not undergone HR was then performed at 37°C where the Cas9 was once again functional. Here we have shown HR as the driving mechanism under temperatures where the nuclease is active and growth is optimal (Mougiakos et al., 2017). The presence of the CRISPR-Cas9 or CRISPR-MAD7 system was not required to detect successful HR due to prototrophic selection on M9 minimal medium. However, where such a selection is not possible, the CRISPR-Cas9 or CRISPR-MAD7 systems act as a powerful counterselection for unedited cells. Additionally, the editing efficiency reported here with pBAC0041 and dDNA (Figure 1) in the presence of IPTG for Cas9 expression induction (91%) is similar to that previously reported in the absence of IPTG (89.2%; Price et al., 2019). Thus, editing rates are decoupled from Cas9 expression levels with leaky P_{grac} promoter activity being sufficient to induce the lethal DSB. Moreover, as the natural competency master regulator ComK activates transcription of the primary component of HR, *recA*, and DNA uptake is single-stranded, yielding a substrate with which RecA can bind, HR can readily proceed at the target site before the DSB taking place (Cheo, Bayles, & Yasbin, 1993; Dubnau, 1999).

Furthermore, a catalytically inactive variant of MAD7 was engineered retaining its ability to bind DNA in the presence of a DNA-targeting gRNA. Our data highlights the importance of testing multiple gRNA sequences when optimizing the MAD7-mediated downregulation of a target gene. The level of downregulation is likely influenced by a combination of factors, such as PAM site sequence, gRNA binding efficiency, GC % of the protospacer, and gRNA secondary structure (Labun et al., 2016; Thyme et al., 2016; Wilson, O'Brien, & Bauer, 2018; Zetsche et al., 2015). Interestingly, the significance of gRNA selection does not seem to be as great when preforming MAD7-mediated editing compared to CRISPRi. The *amyE* +21 targeting PAM site resulted in 98% editing efficiency while only showing a 7.9% decrease of α-amylase activity when preforming CRISPRi with dMAD7. This is likely because the nuclease DSB is a single event whereas CRISPRi requires a stable and continuous interaction between the nuclease and targeted coding sequence for efficient transcriptional downregulation. Multiplexed targeting of dMAD7 to an endogenous and heterologous gene target was exemplified, with no detectable competition between gRNAs on the pool of expressed dMAD7 when compared to strains with only a single gRNA. It is feasible that technologies developed were effector proteins are fused to dCas9 for pathway optimization or in vivo mutagenesis could be adapted for use with dMAD7 further expanding its potential (Bikard et al., 2013; Wang et al., 2018; Yang et al., 2018).

An attractive feature of MAD7 is its freedom to use for industrial R&D and commercial strain construction (provided the final strain does not retain the MAD7 nuclease (Inscripta, 2019b)). This has the capacity to disrupt the slow commercial uptake of genome editing technologies allowing the use of MAD7 in sectors that were put off by licensing and royalty fees associated with for instance Cas9 and Cpf1 CRISPR nucleases, such as large-scale fermentation biotechnology.

ACKNOWLEDGMENTS

The authors also thank Inscripta Inc. for supplying the sequence of MAD7 as an open-source nuclease for the gene-editing community. The authors declare R. C., and F. E.'s affiliation to Ingenza Ltd. The authors also declare that no conflict of interests exists, there are no patents pending for this research, and no products in the development of marketing.

ORCID

Marcus A. Price http://orcid.org/0000-0002-7435-873X
Susan J. Rosser http://orcid.org/0000-0002-2560-6485

REFERENCES

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*, 1911. https://doi.org/10.1038/s41467-018-04252-2

Altenbuchner, J. (2016). Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system. *Applied and Environmental Microbiology*, *82*(17), 5421–5427. https://doi.org/10.1128/AEM.01453-16

Anagnostopoulos, C., & Spizizen, J. (1961). Requirements for transformation in *Bacillus subtilis*. *Journal of Bacteriology*, *81*(5), 741–746.

Bennallack, P. R., Burt, S. R., Heder, M. J., Robison, R. A., & Griffitts, J. S. (2014). Characterization of a novel plasmid-borne thiopeptide gene cluster in *Staphylococcus epidermidis* strain 115. *Journal of Bacteriology*, *196*(24), 4344–4350. https://doi.org/10.1128/JB.02243-14

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., & Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, *41*(15), 7429–7437. https://doi.org/10.1093/nar/gkt520

Bisicchia, P., Botella, E., & Devine, K. M. (2010). Suite of novel vectors for ectopic insertion of GFP, CFP and IYFP transcriptional fusions in single copy at the *amyE* and *bglS* loci in *Bacillus subtilis*. *Plasmid*, *64*(3), 143–149. https://doi.org/10.1016/j.plasmid.2010.06.002

Bryksin, A. V., & Matsumura, I. (2010). Overlap extension PCR cloning: A simple and reliable way to create recombinant plasmids. *Biotechniques*, *48*(6), 463–465. https://doi.org/10.2144/000113418

Burby, P. E., & Simmons, L. A. (2017). MutS2 promotes homologous recombination in *Bacillus subtilis*. *Journal of Bacteriology*, *199*(2), https://doi.org/10.1128/JB.00682-16

Che, A., Knight, T., Canton, B., Kelly, J., & Shetty, R. (2015). Iti Scotland Limited. *United States Patent No. US 8,999,679 B2*. United States of America.

Cheo, D. L., Bayles, K. W., & Yasbin, R. E. (1993). Elucidation of regulatory elements that control damage induction and competence induction of the *Bacillus subtilis* SOS system. *Journal of Bacteriology*, *175*(18), 5907–5915. https://doi.org/10.1128/jb.175.18.5907-5915.1993

Dicarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research*, *41*(7), 4336–4343. https://doi.org/10.1093/nar/gkt135

Dubnau, D. (1999). DNA uptake in bacteria. *Annual Review of Microbiology*, *53*(1), 217–244. https://doi.org/10.1146/annurev.micro.53.1.217

Harwood, C. R., & Cutting, S. M. (1990). *Molecular Biological Methods for Bacillus*. John Wiley & Sons.

Inscripta, I. (2019a). Introducing MAD7 | Inscripta. Retrieved from https://www.inscripta.com/madzymes/

Inscripta, I. (2019b). MADzyme FAQs | Inscripta. Retrieved from https://www.inscripta.com/madzymes/faq/

Jakočiūnas, T., Jensen, M. K., & Keasling, J. D. (2015). CRISPR/Cas9 advances engineering of microbial cell factories. *Metabolic Engineering*, *34*, 44–59. https://doi.org/10.1016/j.ymben.2015.12.003

Jakutyte-Giraitiene, L., & Gasiunas, G. (2016). Design of a CRISPR-Cas system to increase resistance of *Bacillus subtilis* to bacteriophage SPP1. *Journal of Industrial Microbiology & Biotechnology*, *43*(8), 1183–1188.

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. a (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), 233–239. https://doi.org/10.1038/nbt.2508

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA—Guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*, 816–822. https://doi.org/10.1126/science.1225829

Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., & Valen, E. (2016). CHOPCHOP v2: A web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Research*, *44*(W1), W272–W276. https://doi.org/10.1093/nar/gkw398

Liu, D., Huang, C., Guo, J., Zhang, P., Chen, T., Wang, Z., & Zhao, X. (2019). Development and characterization of a CRISPR/Cas9n-based multiplex genome editing system for *Bacillus subtilis*. *Biotechnology for Biofuels*, *12*(1), 197. https://doi.org/10.1186/s13068-019-1537-1

Lu, Z., Yang, S., Yuan, X., Shi, Y., Ouyang, L., Jiang, S., & Zhang, G. (2019). CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in *Bacillus subtilis*. *Nucleic Acids Research*, 13. https://doi.org/10.1093/nar/gkz072

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, *42*(W1), W401–W407. https://doi.org/10.1093/nar/gku410

Mougiakos, I., Bosma, E. F., Weenink, K., Vossen, E., Goijvaerts, K., Van Der Oost, J., & Van Kranenburg, R. (2017). Efficient genome editing of a facultative thermophile using mesophilic spCas9. *ACS Synthetic Biology*, *6*(5), 849–861. https://doi.org/10.1021/acssynbio.6b00339

Peng, F., Wang, X., Sun, Y., Dong, G., Yang, Y., Liu, X., & Bai, Z. (2017). Efficient gene editing in *Corynebacterium glutamicum* using the CRISPR/Cas9 system. *Microbial Cell Factories*, *16*(1), 201. https://doi.org/10.1186/s12934-017-0814-6

Peters, J. M., Colavin, A., Shi, H., Czarny, T. L., Larson, M. H., Wong, S., & Gross, C. A. (2016). A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell*, *165*(6), 1493–1506. https://doi.org/10.1016/j.cell.2016.05.003

Price, M. A., Cruz, R., Baxter, S., Escalettes, F., & Rosser, S. J. (2019). CRISPR-Cas9 *In Situ* engineering of subtilisin E in *Bacillus subtilis*. *PLoS One*, *14*(1), https://doi.org/10.1371/journal.pone.0210121

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Sambrook, J., & Russell, R. W. (2001). Molecular cloning: A laboratory manual. 4th Ed.

Shuman, S., & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nature Reviews Microbiology*, *5*(11), 852–861. https://doi.org/10.1038/nrmicro1768

Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E., & Schier, A. F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nature Communications*, *7*(1), https://doi.org/10.1038/ncomms11750

Wang, Y., Liu, Y., Liu, J., Guo, Y., Fan, L., Ni, X., & Ma, Y. (2018). MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method. *Metabolic Engineering*, *47*, 200–210. https://doi.org/10.1016/j.ymben.2018.02.016

Westbrook, A. W., Moo-Young, M., & Chou, C. P. (2016a). Development of a CRISPR-Cas9 tool kit for comprehensive engineering of *Bacillus subtilis*. *Applied and Environmental Microbiology*, *82*(16), 4876–4895. https://doi.org/10.1128/AEM.01159-16

Westbrook, A. W., Moo-Young, M., & Chou, C. P. (2016b). Development of a CRISPR-Cas9 toolkit for comprehensive engineering of *Bacillus*. *Applied and Environmental Microbiology*, *82*(16), 4876–4895.

Wierson, W. A., Simone, B. W., Warejoncas, Z., Mann, C., Welker, J. M., Gendron, W. A. C., & Essner, J. J. (2019). Expanding the CRISPR toolbox with ErCas12a in zebrafish and human cells. *BioRxiv*, https://doi.org/10.1101/650515

Wilson, L. O. W., O'Brien, A. R., & Bauer, D. C. (2018). The current state and future of CRISPR-Cas9 gRNA design tools. *Frontiers in Pharmacology*, *9*, 749. https://doi.org/10.3389/fphar.2018.00749

Xu, J., Lian, W., Jia, Y., Li, L., & Huang, Z. (2017). Optimized guide RNA structure for genome editing via Cas9. *Oncotarget*, *8*(55), 94166–94171. https://doi.org/10.18632/oncotarget.21607

Yamano, T., Nishimasu, H., Zetsche, B., Ishitani, R., Zhang, F., Correspondence, O. N., & Nureki, O. (2016). Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell*, *165*, 949–962. https://doi.org/10.1016/j.cell.2016.04.003

Yang, L., Zhang, X., Wang, L., Yin, S., Zhu, B., Xie, L., & Li, D. (2018). Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein & Cell*, *9*(9), 814–819. https://doi.org/10.1007/s13238-018-0568-x

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-cas system. *Cell*, *163*(3), 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zhang, K., Duan, X., & Wu, J. (2016). Multigene disruption in undomesticated *Bacillus subtilis* ATCC 6051a using the CRISPR/Cas9 system. *Scientific Reports*, *6*(27943), 1–11.

Zhang, X. Z., Yan, X., Cui, Z. L., Hong, Q., & Li, S. P. (2006). *mazF*, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. *Nucleic Acids Research*, *34*(9), e71. https://doi.org/10.1093/nar/gkl358

Zhu, X., Zhao, D., Qiu, H., Fan, F., Man, S., Bi, C., & Zhang, X. (2017). The CRISPR/Cas9-facilitated multiplex pathway optimization (CFPO) technique and its application to improve the *Escherichia coli* xylose

utilization pathway. *Metabolic Engineering*, *43*, 37–45. https://doi.org/10.1016/j.ymben.2017.08.003

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Price MA, Cruz R, Bryson J, Escalettes F, Rosser SJ. Expanding and understanding the CRISPR toolbox for *Bacillus subtilis* with MAD7 and dMAD7. *Biotechnology and Bioengineering*. 2020;1–12. https://doi.org/10.1002/bit.27312

Bibliography

1. Price, M. A., Cruz, R., Baxter, S., Escalettes, F. & Rosser, S. J. CRISPR-Cas9 *In Situ* engineering of subtilisin E in *Bacillus subtilis*. *PLoS One* **14**, e0210121 (2019).
2. Price, M. A., Cruz, R., Bryson, J., Escalettes, F. & Rosser, S. J. Expanding and understanding the CRISPR toolbox for *Bacillus subtilis* with MAD7 and dMAD7. *Biotechnol. Bioeng.* bit.27312 (2020). doi:10.1002/bit.27312
3. Hohmann, H.-P., van Dijl, J. M., Krishnappa, L. & Prágai, Z. Host Organisms: *Bacillus subtilis*. in *Industrial Biotechnology* 221–297 (2016). doi:10.1002/9783527807796.ch7
4. Gaj, T., Gersbach, C. A. & Barbas III, C. F. ZFN, TALEN and CRISPR/Cas based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2014).
5. Gurumurthy, C. B., Grati, M. & Ohtsuka, M. CRISPR : a versatile tool for both forward and reverse genetics. *Hum. Genet.* (2016). doi:10.1007/s00439-016-1704-4
6. de Souza, P. M. & e Magalhães, P. de O. Application of microbial α-amylase in industry - a review. *Brazilian J. Microbiol.* **41**, 850–861 (2010).
7. Kunst, F. *et al.* The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256 (1997).
8. Harwood, C. R. & Cutting, S. M. *Molecular Biological Methods for Bacillus*. *John Wiley & Sons* (1990).
9. Schallmey, M. *et al.* Developments in the use of *Bacillus* species for industrial production. *J. Microbiol* **50**, 1–17 (2004).
10. Degering, C. *et al.* Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. *Appl. Environ. Microbiol.* **76**, 6370–6376 (2010).
11. BCC Research Staff. *Global Markets for Enzymes in Industrial Applications*. *BCC Research Report Overview* (2018). doi:10.1016/S1351-4180(14)70133-3
12. Saito, N. & Yamamoto, K. Regulatory factors affecting alpha-Amylase production in *Bacillus licheniformis*. *J. Bacteriol.* **121**, 848–856 (1975).
13. Chen, L. *et al.* Induced maize salt tolerance by rhizosphere inoculation of *Bacillus amyloliquefaciens* SQR9. *Physiol. Plant.* **158**, 34–44 (2016).
14. Li, B. *et al.* Increased fermentative adenosine production by gene-targeted *Bacillus subtilis* mutation. *J. Biotechnol.* **298**, 1–4 (2019).
15. Perkins, J. B. *et al.* Genetic engineering of *Bacillus subtilis* for the commercial production of riboflavin. *J. Ind. Microbiol. Biotechnol.* **22**, 8–18 (1999).
16. Anagnostopoulos, C. & Spizizen, J. Requirements for transformation in *Bacillus subtilis*. *J. Bacteriol.* **81**, 741–746 (1961).
17. Zhu, B. & Stülke, J. *Subti*Wiki in 2018: From genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. *Nucleic Acids Res.* **46**, D743–D748 (2018).
18. Rahmer, R., Heravi, K. M. & Altenbuchner, J. Construction of a super-competent *Bacillus subtilis* 168 using the P*mtlA-comKS* inducible cassette. *Front. Microbiol.* **6**, 1431 (2015).
19. Blokesch, M. Natural competence for transformation. *Current Biology* **26**, R1126–R1130 (2016).

20. Hoffmann, K. *et al.* Facilitation of direct conditional knockout of essential genes in *Bacillus licheniformis* DSM13 by comparative genetic analysis and manipulation of genetic competences. *Appl. Environ. Microbiol.* **76**, 5046–5057 (2010).

21. Berka, R. M. *et al.* Microarray analysis of the *Bacillus subtilis* K-state: Genome-wide expression changes dependent on ComK. *Mol. Microbiol.* **43**, 1331–1345 (2002).

22. Inamine, G. S. & Dubnau, D. ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *J. Bacteriol.* **177**, 3045–3051 (1995).

23. Kunst, F. & Rapoport, G. Salt Stress Is an Environmental Signal Affecting Degradative Enzyme Synthesis in Bacillus subtilis. *J. Bacteriol.* **177**, 2403–2407 (1995).

24. Bennallack, P. R., Burt, S. R., Heder, M. J., Robison, R. A. & Griffitts, J. S. Characterization of a Novel Plasmid-Borne Thiopeptide Gene Cluster in *Staphylococcus epidermidis* Strain 115. *J. Bacteriol.* **196**, 4344–4350 (2014).

25. Kidane, D. *et al.* Evidence for different pathways during horizontal gene transfer in competent *Bacillus subtilis* cells. *PLoS Genet.* **5**, 1000630 (2009).

26. Tadesse, S. & Graumann, P. L. DprA/Smf protein localizes at the DNA uptake machinery in competent *Bacillus subtilis* cells. *BMC Microbiol.* **7**, (2007).

27. Zhang, X. Z., Yan, X., Cui, Z. L., Hong, Q. & Li, S. P. *mazF*, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. *Nucleic Acids Res.* **34**, e71 (2006).

28. Fabret, C., Ehrlich, S. D. & Noirot, P. A new mutation delivery system for genome-scale approaches in *Bacillus subtilis*. *Mol. Microbiol.* **46**, 25–36 (2002).

29. Brans, A., Filée, P., Chevigné, A., Claessens, A. & Joris, B. New integrative method to generate *Bacillus subtilis* recombinant strains free of selection markers. *Appl. Environ. Microbiol.* **70**, 7241–7250 (2004).

30. Wang, Y. *et al. Bacillus subtilis* genome editing using ssDNA with short homology regions. *Nucleic Acids Res.* **40**, (2012).

31. Abremski, K., Hoess, R. & Sternberg, N. Studies on the properties of P1 site-specific recombination: Evidence for topologically unlinked products following recombination. *Cell* **32**, 1301–1311 (1983).

32. Albert, H., Dale, E. C., Lee, E. & Ow, D. W. Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. *Plant J.* **7**, 649–659 (1995).

33. Ling Lin Fu *et al.* Protein secretion pathways in *Bacillus subtilis*: Implication for optimization of heterologous protein secretion. *Biotechnol. Adv.* **25**, 1–12 (2007).

34. Zanen, G. *et al.* Proteomic dissection of potential signal recognition particle dependence in protein secretion by *Bacillus subtilis*. *Proteomics* **6**, 3636–3648 (2006).

35. Bolhuis, A. *et al.* SecDF of *Bacillus subtilis*, a molecular siamese twin required for the efficient secretion of proteins. *J. Biol. Chem.* **273**, 21217–21224 (1998).

36. Quesada-Ganuza, A. *et al.* Identification and optimization of PrsA in *Bacillus subtilis* for improved yield of amylase. *Microb. Cell Fact.* **18**, 1–16 (2019).

37. Kouwen, T. R. H. M. & van Dijl, J. M. Interchangeable modules in bacterial thiol-disulfide exchange pathways. *Trends Microbiol.* **17**, 6–12 (2009).

38. Sarvas, M., Harwood, C. R., Bron, S. & Van Dijl, J. M. Post-translocational folding of secretory proteins in Gram-positive bacteria. *Biochim. Biophys. Acta - Mol. Cell Res.* **1694**, 311–327 (2004).

39. Bryan, P. N. Protein engineering of subtilisin. *Biochim. Biophys. Acta* **1543**, 203–222

(2000).

40. Pohl, S. *et al.* Proteomic analysis of *Bacillus subtilis* strains engineered for improved production of heterologous proteins. *Proteomics* **13**, 3298–3308 (2013).

41. Liu, X., Wang, H., Wang, B. & Pan, L. Efficient production of extracellular pullulanase in *Bacillus subtilis* ATCC6051 using the host strain construction and promoter optimization expression system. *Microb. Cell Fact.* **17**, 163 (2018).

42. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).

43. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–822 (2012).

44. Shmakov, S. *et al.* Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems. *Mol Cell* **60**, 385–397 (2015).

45. Gao, P., Yang, H., Rajashankar, K. R., Huang, Z. & Patel, D. J. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* 1–13 (2016). doi:10.1038/cr.2016.88

46. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).

47. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346**, 1258096–1258096 (2014).

48. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. The basic building blocks and evolution of CRISPR-CAS systems. *Biochem. Soc. Trans.* **41**, 1392–400 (2013).

49. Kocak, D. D. *et al.* Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat. Biotechnol.* **37**, 657–666 (2019).

50. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science (80-.).* **348**, 1477–1481 (2015).

51. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233–239 (2013).

52. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).

53. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science (80-.).* **351**, 867–871 (2016).

54. Wang, Y. *et al.* MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method. *Metab. Eng.* **47**, 200–210 (2018).

55. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).

56. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).

57. Eid, A., Alshareef, S. & Mahfouz, M. M. CRISPR base editors: genome editing without double-stranded breaks. *Biochem. J.* **475**, 1955–1964 (2018).

58. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).

59. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).

60. Lu, Z. *et al.* CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in *Bacillus subtilis*. *Nucleic Acids Res.* 13 (2019). doi:10.1093/nar/gkz072

61. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–96 (2013).

62. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).

63. Altenbuchner, J. Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **82**, 5421–5427 (2016).

64. Haeussler, M. & Concordet, J. P. Genome Editing with CRISPR-Cas9: Can It Get Any Better? *Journal of Genetics and Genomics* **43**, 239–250 (2016).

65. Burstein, D. *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).

66. Gao, L. *et al.* Engineered Cpf1 variants with altered PAM specificities increase genome targeting range. *Nat. Biotechnol.* **35**, 789–792 (2017).

67. Inscripta, I. Introducing MAD7 | Inscripta. (2019). Available at: https://www.inscripta.com/madzymes/.

68. Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962 (2016).

69. Safari, F., Zare, K., Negahdaripour, M., Barekati-Mowahed, M. & Ghasemi, Y. CRISPR Cpf1 proteins: Structure, function and implications for genome editing. *Cell and Bioscience* **9**, (2019).

70. Inscripta, I. MADzyme FAQs | Inscripta. (2019). Available at: https://www.inscripta.com/madzymes/faq/.

71. Wierson, W. A. *et al.* Expanding the CRISPR Toolbox with ErCas12a in Zebrafish and Human Cells. *bioRxiv* (2019). doi:10.1101/650515

72. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5–15 (2016).

73. Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493–1506 (2016).

74. Zhang, X. *et al.* Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discov.* **3**, (2017).

75. Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R. & van der Oost, J. Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends Biotechnol.* **34**, 575–587 (2016).

76. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat Methods* **10**, 957–963 (2013).

77. Jensen, E. D. *et al.* Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. *Microb. Cell Fact.* **16**, (2017).

78. Dong, C. Programming Bacterial Gene Expression Using Synthetic CRISPR-Cas Transcriptional Regulators. (2019).

79. Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).

80. Gaudelli, N. M. *et al.* Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

81. Lenhart, J. S., Schroeder, J. W., Walsh, B. W. & Simmons, L. a. DNA Repair and Genome Maintenance in *Bacillus subtilis*. *Microbiol. Mol. Biol. Rev.* **76**, 530–564 (2012).

82. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.

Nature **533**, 420–424 (2016).

83. Zheng, K. *et al.* Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. *Commun. Biol.* **1**, 32 (2018).

84. Wang, Y. *et al.* Expanding targeting scope, editing window, and base transition capability of base editing in *Corynebacterium glutamicum*. *Biotechnol. Bioeng.* **116**, 3016–3029 (2019).

85. Karathanasis, E. & Wilson, T. E. Enhancement of *Saccharomyces cerevisiae* end-joining efficiency by cell growth stage but not by impairment of recombination. *Genetics* **161**, 1015–1027 (2002).

86. Daley, J. M., Palmbos, P. L., Wu, D. & Wilson, T. E. Nonhomologous end joining in yeast. *Annu. Rev. Genet.* **39**, 431–451 (2005).

87. Kidane, D. & Graumann, P. L. Intracellular protein and DNA dynamics in competent *Bacillus subtilis* cells. *Cell* **122**, 73–84 (2005).

88. Rösch, T. C. *et al.* Single molecule tracking reveals spatio-temporal dynamics of bacterial DNA repair centres. *Sci. Rep.* **8**, (2018).

89. Cardenas, P. P. *et al.* Polynucleotide phosphorylase exonuclease and polymerase activities on single-stranded DNA ends are modulated by RecN, SsbA and RecA proteins. *Nucleic Acids Res.* **39**, 9250–9261 (2011).

90. Maffeo, C. & Aksimentiev, A. Molecular mechanism of DNA association with single-stranded DNA binding protein. *Nucleic Acids Res.* **45**, 12125–12139 (2017).

91. Tang, Q. *et al.* ATP-dependent conformational change in ABC-ATPase RecF serves as a switch in DNA repair. *Sci. Rep.* **8**, (2018).

92. Murray, H. & Koh, A. Multiple Regulatory Systems Coordinate DNA Replication with Cell Growth in *Bacillus subtilis*. *PLoS Genet.* **10**, 1004731 (2014).

93. Torres, R., Romero, H., Rodríguez-Cerrato, V. & Alonso, J. C. Interplay between *Bacillus subtilis* RecD2 and the RecG or RuvAB helicase in recombinational repair. *DNA Repair (Amst).* **55**, 40–46 (2017).

94. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. *Annu. Rev. Genet.* **38**, 233–271 (2004).

95. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E924-32 (2014).

96. Wang, S. T. *et al.* The forespore line of gene expression in *Bacillus subtilis*. *J. Mol. Biol.* **358**, 16–37 (2006).

97. De Ory, A., Zafra, O. & De Vega, M. Efficient processing of abasic sites by bacterial nonhomologous end-joining Ku proteins. *Nucleic Acids Res.* **42**, 13082–13095 (2014).

98. Haldenwang, W. G. The sigma factors of *Bacillus subtilis*. *Microbiol.Rev.* **59**, 1–30 (1995).

99. Ayora, S. *et al.* Double-strand break repair in bacteria: A view from *Bacillus subtilis*. *FEMS Microbiol. Rev.* **35**, 1055–1081 (2011).

100. Alonso, J. C. *et al.* Early steps of double-strand break repair in *Bacillus subtilis*. *DNA Repair (Amst).* **12**, 162–176 (2013).

101. de Vega, M. The Minimal *Bacillus subtilis* Nonhomologous End Joining Repair Machinery. *PLoS One* **8**, (2013).

102. Toymentseva, A. A. & Altenbuchner, J. New CRISPR-Cas9 vectors for genetic modifications of *Bacillus* species. *FEMS Microbiol. Lett.* **366**, (2019).

103. Zheng, X., Li, S. Y., Zhao, G. P. & Wang, J. An efficient system for deletion of large DNA fragments in *Escherichia coli* via introduction of both Cas9 and the non-homologous end joining system from *Mycobacterium smegmatis*. *Biochem. Biophys. Res. Commun.* **485**, 768–774 (2017).

104. Oh, J. H. & Van Pijkeren, J. P. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res.* **42**, (2014).

105. Jakutyte-Giraitiene, L. & Gasiunas, G. Design of a CRISPR-Cas system to increase resistance of *Bacillus subtilis* to bacteriophage SPP1. *J. Ind. Microbiol. Biotechnol.* **43**, 1183–1188 (2016).

106. Westbrook, A. W., Moo-Young, M. & Chou, C. P. Development of a CRISPR-Cas9 tool kit for comprehensive engineering of *Bacillus subtilis*. *Appl. Environ. Microbiol.* **82**, 4876–4895 (2016).

107. Westbrook, A. W., Ren, X., Oh, J., Moo-Young, M. & Chou, C. P. Metabolic engineering to enhance heterologous production of hyaluronic acid in *Bacillus subtilis*. *Metab. Eng.* **47**, 401–413 (2018).

108. Westbrook, A. W., Ren, X., Moo-Young, M. & Chou, C. P. Engineering of cell membrane to enhance heterologous production of hyaluronic acid in *Bacillus subtilis*. *Biotechnol. Bioeng.* **115**, 216–231 (2018).

109. Watzlawick, H. & Altenbuchner, J. Multiple integration of the gene *ganA* into the *Bacillus subtilis* chromosome for enhanced β-galactosidase production using the CRISPR/Cas9 system. *AMB Express* **9**, (2019).

110. Schilling, T. *et al.* A CRISPR-Cas9-Based Toolkit for Fast and Precise In Vivo Genetic Engineering of *Bacillus subtilis* Phages. *Viruses* **10**, (2018).

111. Yi, Y., Li, Z., Song, C. & Kuipers, O. P. Exploring plant-microbe interactions of the rhizobacteria *Bacillus subtilis* and *Bacillus mycoides* by use of the CRISPR-Cas9 system. *Environ. Microbiol.* **20**, 4245–4260 (2018).

112. Zhang, K., Duan, X. & Wu, J. Multigene disruption in undomesticated *Bacillus subtilis* ATCC 6051a using the CRISPR/Cas9 system. *Sci. Rep.* **6**, 1–11 (2016).

113. Burby, P. E. & Simmons, L. A. MutS2 promotes homologous recombination in *Bacillus subtilis*. *J. Bacteriol.* **199**, (2017).

114. So, Y. *et al.* A Highly efficient CRISPR-Cas9-mediated large genomic deletion in *Bacillus subtilis*. *Front. Microbiol.* **8**, 1167 (2017).

115. Li, K., Cai, D., Wang, Z., He, Z. & Chen, S. Development of an Efficient Genome Editing Tool in *Bacillus licheniformis* Using CRISPR-Cas9 Nickase. *Appl. Enviromental Microbiol.* **84**, (2018).

116. Liu, D. *et al.* Development and characterization of a CRISPR/Cas9n-based multiplex genome editing system for *Bacillus subtilis*. *Biotechnol. Biofuels* **12**, 197 (2019).

117. Mougiakos, I. *et al.* Efficient Genome Editing of a Facultative Thermophile Using Mesophilic spCas9. *ACS Synth. Biol.* **6**, 849–861 (2017).

118. Guiziou, S. *et al.* A part toolbox to tune genetic expression in *Bacillus subtilis*. *Nucleic Acids Res.* **44**, 7495–7508 (2016).

119. Sambrook, J. & Russell, R. W. *Molecular Cloning: A laboratory manual*. (2001).

120. Bryksin, A. V & Matsumura, I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* **48**, 463–5 (2010).

121. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murray, N. E. Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res.* **42**, 3–19 (2014).

122. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred

kilobases. *Nat. Methods* **6**, 343–345 (2009).

123. Che, A., Knight, T., Canton, B., Kelly, J. & Shetty, R. United States Patent No. US 8,999,679 B2. 1–54 (2015).

124. Bisicchia, P., Botella, E. & Devine, K. M. Suite of novel vectors for ectopic insertion of GFP, CFP and IYFP transcriptional fusions in single copy at the *amyE* and *bglS* loci in *Bacillus subtilis*. *Plasmid* **64**, 143–149 (2010).

125. Liu, H. *et al.* CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation: Fig. 1. *Bioinformatics* **31**, 3676–3678 (2015).

126. Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44**, W272–W276 (2016).

127. Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407 (2014).

128. Cupp-Enyard, C. Sigma's Non-specific Protease Activity Assay - Casein as a Substrate. *J. Vis. Exp.* (2008). doi:10.3791/899

129. Jain, S. C., Shinde, U., Li, Y., Inouye, M. & Berman, H. M. The crystal structure of an autoprocessed Ser221Cys-subtilisin E-propeptide complex at 2.0 å resolution. *J. Mol. Biol.* **284**, 137–144 (1998).

130. Shirai, T. *et al.* High-resolution crystal structure of M-protease: phylogeny aided analysis of the high-alkaline adaptation mechanism. *Protein Eng.* **10**, 627–634 (1997).

131. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).

132. Giuliani, S. E., Frank, A. M. & Collart, F. R. Functional Assignment of Solute-Binding Proteins of ABC Transporters Using a Fluorescence-Based Thermal Shift Assay. *Biochemistry* **47**, 13974–13984 (2008).

133. Bustin, S. A. *et al.* Clinical Chemistry. *Clin. Chem.* **48**, 1178–1185 (2009).

134. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, 71–74 (2007).

135. Westbrook, A. W., Moo-Young, M. & Chou, C. P. Development of a CRISPR-Cas9 toolkit for comprehensive engineering of *Bacillus*. *Appl. Environ. Microbiol.* **82(16)**, 4876–4895 (2016).

136. Lilly, J. & Camps, M. Mechanisms of Theta Plasmid Replication. *Microbiol. Spectr.* **3**, (2015).

137. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

138. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

139. Radeck, J. *et al.* The *Bacillus* BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *J. Biol. Eng.* **7**, 29 (2013).

140. Sojka, L. *et al.* Rapid changes in gene expression: DNA determinants of promoter regulation by the concentration of the transcription initiating NTP in *Bacillus subtilis*. *Nucleic Acids Res.* **39**, 4598–4611 (2011).

141. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nature Communications* **9**, (2018).

142. Jiang, Y. *et al.* CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. *Nat. Commun.* **8**, 15179 (2017).

143. Cho, S. *et al.* High-Level dCas9 Expression Induces Abnormal Cell Morphology in *Escherichia coli*. *ACS Synth. Biol.* **7**, 1085–1094 (2018).

144. Jensen, C. L., Stephenson, K., Jørgensen, S. T. & Harwood, C. Cell-associated degradation affects the yield of secreted engineered and heterologous proteins in the *Bacillus subtilis* expression system. *Microbiology* **146**, 2583–2594 (2015).

145. Westers, L. *et al.* Genetic or chemical protease inhibition causes significant changes in the *Bacillus subtilis* exoproteome. *Proteomics* **8**, 2704–2713 (2008).

146. Huang, G. L., Gosschalk, J. E., Kim, Y. S., Ogorzalek Loo, R. R. & Clubb, R. T. Stabilizing displayed proteins on vegetative *Bacillus subtilis* cells. *Appl. Microbiol. Biotechnol.* **102**, 6547–6565 (2018).

147. Fujita, M. & Losick, R. The master regulator for entry into sporulation in *Bacillus subtilis* becomes a cell-specific transcription factor after asymmetric division. *Genes Dev.* **17**, 1166–1174 (2003).

148. Cheo, D. L., Bayles, K. W. & Yasbin, R. E. Elucidation of regulatory elements that control damage induction and competence induction of the *Bacillus subtilis* SOS system. *J. Bacteriol.* **175**, 5907–5915 (1993).

149. Dubnau, D. DNA Uptake in Bacteria. *Annu. Rev. Microbiol.* **53**, 217–244 (1999).

150. Bednarska, N. G., Schymkowitz, J., Rousseau, F. & Van Eldere, J. Protein aggregation in bacteria: The thin boundary between functionality and toxicity. *Microbiology* **159**, 1795–1806 (2013).

151. Haddar, A. *et al.* Two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: Purification, characterization and potential application as a laundry detergent additive. *Bioresour. Technol.* **100**, 3366–3373 (2009).

152. Sharma, K. M., Kumar, R., Panwar, S. & Kumar, A. Microbial alkaline proteases: Optimization of production parameters and their properties. *Journal of Genetic Engineering and Biotechnology* **15**, 115–126 (2017).

153. Kobayashi, T. *et al.* Contribution of a salt bridge triad to the thermostability of a highly alkaline protease from an alkaliphilic *Bacillus* strain. *World J. Microbiol. Biotechnol.* **21**, 961–967 (2005).

154. Britton, R. A. *et al.* Genome-wide analysis of the stationary-phase sigma factor (sigma-H) regulon of *Bacillus subtilis*. *J. Bacteriol.* **184**, 4881–90 (2002).

155. Rocha, E. P. C., Danchin, A. & Viari, A. Translation in *Bacillus subtilis*: Roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Res.* **27**, 3567–3576 (1999).

156. Vellanoweth, R. L. & Rabinowitz, J. C. The influence of ribosome-binding-site elements on translational efficiency in Bacillus subtilis and Escherichia coli in vivo. *Mol. Microbiol.* **6**, 1105–1114 (1992).

157. Kohlstaedt, M., Von Der Hocht, I., Hilbers, F., Thielmann, Y. & Michel, H. Development of a Thermofluor assay for stability determination of membrane proteins using the Na+/H+ antiporter NhaA and cytochrome c oxidase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 1112–1122 (2015).

158. Erwin, C. R., Barnett, B. L., Oliver, J. D. & Sullivan, J. F. Effects of engineered salt bridges on the stability of subtilisin BPN′. *Protein Eng. Des. Sel.* **4**, 87–97 (1990).

159. Titok, M. A. *et al. Bacillus subtilis* soil isolates: Plasmid replicon analysis and construction of a new theta-replicating vector. *Plasmid* **49**, 53–62 (2003).

160. Yang, L. *et al.* Engineering and optimising deaminase fusions for genome editing. *Nature communications* **8**, 16169 (2017).

161. Weiss, A. *et al.* The ω subunit governs RNA polymerase stability and transcriptional specificity in *Staphylococcus aureus*. *J. Bacteriol.* **199**, (2017).

162. Weiss, A. & Shaw, L. N. Small things considered: The small accessory subunits of RNA polymerase in Gram-positive bacteria. *FEMS Microbiology Reviews* **39**, 541–554 (2015).

163. Dove, S. L. & Hochschild, A. Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. *Genes Dev.* **12**, 745–54 (1998).

164. Mencia, M., Monsalve, M., Rojo, F. & Salas, M. Transcription activation by phage phi29 protein p4 is mediated by interaction with the alpha subunit of *Bacillus subtilis* RNA polymerase. *Proc. Natl. Acad. Sci.* **93**, 6616–6620 (1996).

165. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). in *Gene* **173**, 33–38 (Elsevier B.V., 1996).

166. Michna, R. H., Zhu, B., Mäder, U. & Stülke, J. SubtiWiki 2.0 - An integrated database for the model organism *Bacillus subtilis*. *Nucleic Acids Res.* **44**, D654–D662 (2015).

167. Song, Y. *et al.* Promoter screening from *Bacillus subtilis* in various conditions hunting for synthetic biology and industrial applications. *PLoS One* **11**, 1–18 (2016).

168. Dong, C., Fontana, J., Patel, A., Carothers, J. M. & Zalatan, J. G. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat. Commun.* **9**, (2018).

169. Shah, I. M. & Wolf, R. E. Novel protein-protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase α subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent. *J. Mol. Biol.* **343**, 513–532 (2004).

170. Liu, Y., Wan, X. & Wang, B. Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nat. Commun.* **10**, (2019).

171. Fernández, S. & Alonso, J. C. *Bacillus subtilis* sequence-independent DNA-binding and DNA-bending protein Hbsu negatively controls its own synthesis. *Gene* **231**, 187–193 (1999).

172. Li, L. *et al.* CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in *Streptomyces*. *Appl. Environ. Microbiol.* **84**, 827–845 (2018).

173. Xu, J., Lian, W., Jia, Y., Li, L. & Huang, Z. Optimized guide RNA structure for genome editing via Cas9. *Oncotarget* **8**, 94166–94171 (2017).

174. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat. Commun.* **7**, 11750 (2016).

175. Singh, D. *et al.* Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc. Natl. Acad. Sci.* **115**, 5444–5449 (2018).

176. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).

177. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).

178. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. & Kondo, A. Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **3**, 423–429 (2018).

179. Jiang, F. & Doudna, J. A. CRISPR – Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–531 (2017).

180. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10098–10103 (2011).

181. Fonfara, I., Richter, H., BratoviÄ, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).

182. Anzalone, A. V *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. doi:10.1038/s41586-019-1711-4